

B4

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
28 February 2002 (28.02.2002)

PCT

(10) International Publication Number  
**WO 02/16597 A2**

(51) International Patent Classification<sup>7</sup>: **C12N 15/00**

(21) International Application Number: PCT/US01/26365

(22) International Filing Date: 22 August 2001 (22.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/227,429 23 August 2000 (23.08.2000) US  
60/231,370 8 September 2000 (08.09.2000) US  
60/233,212 15 September 2000 (15.09.2000) US  
60/236,885 29 September 2000 (29.09.2000) US

(71) Applicant (for all designated States except US): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **GRIFFIN, Jennifer**, A. [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **PATTERSON, Chandra** [US/US]; 490 Sherwood Way, #1, Menlo Park, CA 94025 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street, #205, San Leandro, CA 94577 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **DING, Li**

[CN/US]; 3353 Alma Street, #146, Palo Alto, CA 94306 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). **DAS, Debopriya** [IN/US]; 1179 Bonita Avenue, #3, Mountain View, CA 94040 (US). **THORNTON, Michael** [US/US]; 9 Medway Road, Woodside, CA 94062 (US). **LAL, Preeti** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/16597 A2

(54) Title: LIPID METABOLISM ENZYMES

(57) Abstract: The invention provides human lipid metabolism enzymes (LMM) and polynucleotides which identify and encode LMM. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of LMM.

## LIPID METABOLISM ENZYMES

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of lipid metabolism enzymes  
5 and to the use of these sequences in the diagnosis, treatment, and prevention of cancer, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, and cardiovascular disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of lipid metabolism enzymes.

### BACKGROUND OF THE INVENTION

Lipids are water-insoluble, oily or greasy substances that are soluble in nonpolar solvents such as chloroform or ether. Neutral fats (triacylglycerols) serve as major fuels and energy stores. Polar lipids, such as phospholipids, sphingolipids, glycolipids, and cholesterol, are key structural components of cell membranes. (Lipid metabolism is reviewed in Stryer, L. (1995) Biochemistry, W.H. Freeman  
15 and Company, New York NY; Lehninger, A. (1982) Principles of Biochemistry, Worth Publishers, Inc. New York NY; and ExPASy "Biochemical Pathways" index of Boehringer Mannheim World Wide Web site, "<http://www.expasy.ch/cgi-bin/search-biochem-index>".)

Fatty acids are long-chain organic acids with a single carboxyl group and a long non-polar hydrocarbon tail. Long-chain fatty acids are essential components of glycolipids, phospholipids, and  
20 cholesterol, which are building blocks for biological membranes, and of triglycerides, which are biological fuel molecules. Long-chain fatty acids are also substrates for eicosanoid production, and are important in the functional modification of certain complex carbohydrates and proteins. 16-carbon and 18-carbon fatty acids are the most common. Fatty acid synthesis occurs in the cytoplasm. In the first step, acetyl-Coenzyme A (CoA) carboxylase (ACC) synthesizes malonyl-CoA from acetyl-CoA and  
25 bicarbonate. The enzymes which catalyze the remaining reactions are covalently linked into a single polypeptide chain, referred to as the multifunctional enzyme fatty acid synthase (FAS). FAS catalyzes the synthesis of palmitate from acetyl-CoA and malonyl-CoA. FAS contains acetyl transferase, malonyl transferase,  $\beta$ -ketoacetyl synthase, acyl carrier protein,  $\beta$ -ketoacyl reductase, dehydratase, enoyl reductase, and thioesterase activities. The final product of the FAS reaction is the 16-carbon  
30 fatty acid palmitate. Further elongation, as well as unsaturation, of palmitate by accessory enzymes of the ER produces the variety of long chain fatty acids required by the individual cell. These enzymes include a NADH-cytochrome  $b_5$  reductase, cytochrome  $b_5$ , and a desaturase.

Triacylglycerols, also known as triglycerides and neutral fats, are major energy stores in

animals. Triacylglycerols are esters of glycerol with three fatty acid chains. Glycerol-3-phosphate is produced from dihydroxyacetone phosphate by the enzyme glycerol phosphate dehydrogenase or from glycerol by glycerol kinase. Fatty acid-CoA's are produced from fatty acids by fatty acyl-CoA synthetases. Glycerol-3-phosphate is acylated with two fatty acyl-CoA's by the enzyme glycerol phosphate acyltransferase to give phosphatidate. Phosphatidate phosphatase converts phosphatidate to diacylglycerol, which is subsequently acylated to a triacylglycerol by the enzyme diglyceride acyltransferase. Phosphatidate phosphatase and diglyceride acyltransferase form a triacylglycerol synthetase complex bound to the ER membrane.

A major class of phospholipids are the phosphoglycerides, which are composed of a glycerol backbone, two fatty acid chains, and a phosphorylated alcohol. Phosphoglycerides are components of cell membranes. Principal phosphoglycerides are phosphatidyl choline, phosphatidyl ethanolamine, phosphatidyl serine, phosphatidyl inositol, and diphosphatidyl glycerol. Many enzymes involved in phosphoglyceride synthesis are associated with membranes (Meyers, R.A. (1995) Molecular Biology and Biotechnology, VCH Publishers Inc., New York NY, pp. 494-501). Phosphatidate is converted to CDP-diacylglycerol by the enzyme phosphatidate cytidyltransferase (ExPASy ENZYME EC 2.7.7.41). Transfer of the diacylglycerol group from CDP-diacylglycerol to serine to yield phosphatidyl serine, or to inositol to yield phosphatidyl inositol, is catalyzed by the enzymes CDP-diacylglycerol-serine O-phosphatidyltransferase and CDP-diacylglycerol-inositol 3-phosphatidyltransferase, respectively (ExPASy ENZYME EC 2.7.8.8; ExPASy ENZYME EC 2.7.8.11). The enzyme phosphatidyl serine decarboxylase catalyzes the conversion of phosphatidyl serine to phosphatidyl ethanolamine, using a pyruvate cofactor (Voelker, D.R. (1997) *Biochim. Biophys. Acta* 1348:236-244). Phosphatidyl choline is formed using diet-derived choline by the reaction of CDP-choline with 1,2-diacylglycerol, catalyzed by diacylglycerol cholinephosphotransferase (ExPASy ENZYME 2.7.8.2).

Cholesterol, composed of four fused hydrocarbon rings with an alcohol at one end, moderates the fluidity of membranes in which it is incorporated. In addition, cholesterol is used in the synthesis of steroid hormones such as cortisol, progesterone, estrogen, and testosterone. Bile salts derived from cholesterol facilitate the digestion of lipids. Cholesterol in the skin forms a barrier that prevents excess water evaporation from the body. Farnesyl and geranylgeranyl groups, which are derived from cholesterol biosynthesis intermediates, are post-translationally added to signal transduction proteins such as Ras and protein-targeting proteins such as Rab. These modifications are important for the activities of these proteins (Guyton, A.C. (1991) Textbook of Medical Physiology, W.B. Saunders Company, Philadelphia PA, pp. 760-763; Stryer, supra, pp. 279-280, 691-702, 934). Mammals obtain

cholesterol derived from both de novo biosynthesis and the diet.

Sphingolipids are an important class of membrane lipids that contain sphingosine, a long chain amino alcohol. They are composed of one long-chain fatty acid, one polar head alcohol, and sphingosine or sphingosine derivatives. The three classes of sphingolipids are sphingomyelins, 5 cerebroside, and gangliosides. Sphingomyelins, which contain phosphocholine or phosphoethanolamine as their head group, are abundant in the myelin sheath surrounding nerve cells. Galactocerebrosides, which contain a glucose or galactose head group, are characteristic of the brain. Other cerebroside are found in non-neural tissues. Gangliosides, whose head groups contain multiple sugar units, are abundant in the brain, but are also found in non-neural tissues.

10 Eicosanoids, including prostaglandins, prostacyclin, thromboxanes, and leukotrienes, are 20-carbon molecules derived from fatty acids. Eicosanoids are signaling molecules which have roles in pain, fever, and inflammation. The precursor of all eicosanoids is arachidonate, which is generated from phospholipids by phospholipase A<sub>2</sub> and from diacylglycerols by diacylglycerol lipase. Leukotrienes are produced from arachidonate by the action of lipoxygenases.

15 Within cells, fatty acids are transported by cytoplasmic fatty acid binding proteins (Online Mendelian Inheritance in Man (OMIM) \*134650 Fatty Acid-Binding Protein 1, Liver; FABP1). Diazepam binding inhibitor (DBI), also known as endozepine and acyl CoA-binding protein, is an endogenous  $\gamma$ -aminobutyric acid (GABA) receptor ligand which is thought to down-regulate the effects of GABA. DBI binds medium- and long-chain acyl-CoA esters with very high affinity and 20 may function as an intracellular carrier of acyl-CoA esters (OMIM \*125950 Diazepam Binding Inhibitor; DBI; PROSITE PDOC00686 Acyl-CoA-binding protein signature).

Fat stored in liver and adipose triglycerides may be released by hydrolysis and transported in the blood. Free fatty acids are transported in the blood by albumin. Triacylglycerols and cholesterol esters in the blood are transported in lipoprotein particles. The particles consist of a core of 25 hydrophobic lipids surrounded by a shell of polar lipids and apolipoproteins. The protein components serve in the solubilization of hydrophobic lipids and also contain cell-targeting signals. Lipoproteins include chylomicrons, chylomicron remnants, very-low-density lipoproteins (VLDL), intermediate-density lipoproteins (IDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL). There is a strong inverse correlation between the levels of plasma HDL and risk of premature 30 coronary heart disease.

Mitochondrial and peroxisomal beta-oxidation enzymes degrade saturated and unsaturated fatty acids by sequential removal of two-carbon units from CoA-activated fatty acids. The main beta-oxidation pathway degrades both saturated and unsaturated fatty acids while the auxiliary pathway

performs additional steps required for the degradation of unsaturated fatty acids. The pathways of mitochondrial and peroxisomal beta-oxidation use similar enzymes, but have different substrate specificities and functions. Mitochondria oxidize short-, medium-, and long-chain fatty acids to produce energy for cells. Mitochondrial beta-oxidation is a major energy source for cardiac and skeletal muscle. In liver, it provides ketone bodies to the peripheral circulation when glucose levels are low as in starvation, endurance exercise, and diabetes (Eaton, S. et al. (1996) *Biochem. J.* 320:345-357). Peroxisomes oxidize medium-, long-, and very-long-chain fatty acids, dicarboxylic fatty acids, branched fatty acids, prostaglandins, xenobiotics, and bile acid intermediates. The chief roles of peroxisomal beta-oxidation are to shorten toxic lipophilic carboxylic acids to facilitate their excretion and to shorten very-long-chain fatty acids prior to mitochondrial beta-oxidation (Mannaerts, G.P. and P.P. Van Veldhoven (1993) *Biochimie* 75:147-158). Enzymes involved in beta-oxidation include acyl CoA synthetase, carnitine acyltransferase, acyl CoA dehydrogenases, enoyl CoA hydratases, L-3-hydroxyacyl CoA dehydrogenase,  $\beta$ -ketothiolase, 2,4-dienoyl CoA reductase, and isomerase.

Three classes of lipid metabolism enzymes are discussed in further detail. The three classes are lipases, phospholipases and lipoxygenases.

#### Lipases

Triglycerides are hydrolyzed to fatty acids and glycerol by lipases. Adipocytes contain lipases that break down stored triacylglycerols, releasing fatty acids for export to other tissues where they are required as fuel. Lipases are widely distributed in animals, plants, and prokaryotes. Triglyceride lipases (ExPASy ENZYME EC 3.1.1.3), also known as triacylglycerol lipases and tributyrases, hydrolyze the ester bond of triglycerides. In higher vertebrates there are at least three tissue-specific isozymes including gastric, hepatic, and pancreatic lipases. These three types of lipases are structurally closely related to each other as well as to lipoprotein lipase. The most conserved region in gastric, hepatic, and pancreatic lipases is centered around a serine residue which is also present in lipases of prokaryotic origin. Mutation in the serine residue renders the enzymes inactive. Gastric, hepatic, and pancreatic lipases hydrolyze lipoprotein triglycerides and phospholipids. Gastric lipases in the intestine aid in the digestion and absorption of dietary fats. Hepatic lipases are bound to and act at the endothelial surfaces of hepatic tissues. Hepatic lipases also play a major role in the regulation of plasma lipids. Pancreatic lipase requires a small protein cofactor, colipase, for efficient dietary lipid hydrolysis. Colipase binds to the C-terminal, non-catalytic domain of lipase, thereby stabilizing an active conformation and considerably increasing the overall hydrophobic binding site. Deficiencies of these enzymes have been identified in man, and all are associated with pathologic levels of circulating lipoprotein particles (Gargouri, Y. et al. (1989) *Biochim. Biophys. Acta* 1006:255-271; Connelly, P.W.

(1999) Clin. Chim. Acta 286:243-255; van Tilbeurgh, H. et al. (1999) Biochim Biophys Acta 1441:173-184).

Lipoprotein lipases (ExPASy ENZYME EC 3.1.1.34), also known as clearing factor lipases, diglyceride lipases, or diacylglycerol lipases, hydrolyze triglycerides and phospholipids present in circulating plasma lipoproteins, including chylomicrons, very low and intermediate density lipoproteins, and high-density lipoproteins (HDL). Together with pancreatic and hepatic lipases, lipoprotein lipases (LPL) share a high degree of primary sequence homology. Both lipoprotein lipases and hepatic lipases are anchored to the capillary endothelium via glycosaminoglycans and can be released by intravenous administration of heparin. LPLs are primarily synthesized by adipocytes, muscle cells, and macrophages. Catalytic activities of LPLs are activated by apolipoprotein C-II and are inhibited by high ionic strength conditions such as 1 M NaCl. LPL deficiencies in humans contribute to metabolic diseases such as hypertriglyceridemia, HDL2 deficiency, and obesity (Jackson, R.L. (1983) in The Enzymes (Boyer, P.D., ed.) Vol. XVI, pp. 141-186, Academic Press, New York NY; Eckel, R.H. (1989) New Engl. J. Med. 320:1060-1068).

#### 15 Phospholipases

Phospholipases, a group of enzymes that catalyze the hydrolysis of membrane phospholipids, are classified according to the bond cleaved in a phospholipid. They are classified into PLA1, PLA2, PLB, PLC, and PLD families. Phospholipases are involved in many inflammatory reactions by making arachidonate available for eicosanoid biosynthesis. More specifically, arachidonic acid is processed into bioactive lipid mediators of inflammation such as lyso-platelet-activating factor and eicosanoids. The synthesis of arachidonic acid from membrane phospholipids is the rate-limiting step in the biosynthesis of the four major classes of eicosanoids (prostaglandins, prostacyclins, thromboxanes and leukotrienes) which are involved in pain, fever, and inflammation (Kaiser, E. et al. (1990) Clin. Biochem. 23:349-370). Furthermore, leukotriene-B4 is known to function in a feedback loop which further increases PLA2 activity (Wijkander, J. et al. (1995) J. Biol. Chem. 270:26543-26549).

The secretory phospholipase A<sub>2</sub> (PLA2) superfamily comprises a number of heterogeneous enzymes whose common feature is to hydrolyze the sn-2 fatty acid acyl ester bond of phosphoglycerides. Hydrolysis of the glycerophospholipids releases free fatty acids and lysophospholipids. PLA2 activity generates precursors for the biosynthesis of biologically active lipids, hydroxy fatty acids, and platelet-activating factor. PLA2s were first described as components of snake venoms, and were later characterized in numerous species. PLA2s have traditionally been classified into several major groups and subgroups based on their amino acid sequences, divalent

cation requirements, and location of disulfide bonds. The PLA2s of Groups I, II, and III consist of low molecular weight, secreted,  $\text{Ca}^{2+}$ -dependent proteins. Group IV PLA2s are primarily 85-kDa,  $\text{Ca}^{2+}$ -dependent cytosolic phospholipases. Finally, a number of  $\text{Ca}^{2+}$ -independent PLA2s have been described, which comprise Group V (Davidson, F.F. and E.A. Dennis (1990) *J. Mol. Evol.* 31:228-238; and Dennis, E.F. (1994) *J. Biol. Chem.* 269:13057-13060).

The first PLA2s to be extensively characterized were the Group I, II, and III PLA2s found in snake and bee venoms. These venom PLA2s share many features with mammalian PLA2s including a common catalytic mechanism, the same  $\text{Ca}^{2+}$  requirement, and conserved primary and tertiary structures. In addition to their role in the digestion of prey, the venom PLA2s display neurotoxic, myotoxic, anticoagulant, and proinflammatory effects in mammalian tissues. This diversity of pathophysiological effects is due to the presence of specific, high affinity receptors for these enzymes on various cells and tissues (Lambeau, G. et al. (1995) *J. Biol. Chem.* 270:5534-5540).

PLA2s from Groups I, IIA, IIC, and V have been described in mammalian and avian cells, and were originally characterized by tissue distribution, although the distinction is no longer absolute. Thus, Group I PLA2s were found in the pancreas, Group IIA and IIC were derived from inflammation-associated tissues (e.g., the synovium), and Group V were from cardiac tissue. The pancreatic PLA2s function in the digestion of dietary lipids and have been proposed to play a role in cell proliferation, smooth muscle contraction, and acute lung injury. The Group II inflammatory PLA2s are potent mediators of inflammatory processes and are highly expressed in serum and synovial fluids of patients with inflammatory disorders. These Group II PLA2s are found in most human cell types assayed and are expressed in diverse pathological processes such as septic shock, intestinal cancers, rheumatoid arthritis, and epidermal hyperplasia. A Group V PLA2 has been cloned from brain tissue and is strongly expressed in heart tissue. A human PLA2 was recently cloned from fetal lung, and based on its structural properties, appears to be the first member of a new group of mammalian PLA2s, referred to as Group X. Other PLA2s have been cloned from various human tissues and cell lines, suggesting a large diversity of PLA2s (Chen, J. et al. (1994) *J. Biol. Chem.* 269:2365-2368; Kennedy, B.P. et al. (1995) *J. Biol. Chem.* 270: 22378-22385; Komada, M. et al. (1990) *Biochem. Biophys. Res. Commun.* 168:1059-1065; Cupillard, L. et al. (1997) *J. Biol. Chem.* 272:15745-15752; and Nalefski, E.A. et al. (1994) *J. Biol. Chem.* 269:18239-18249).

Phospholipases B (PLB) (ExPASy ENZYME EC 3.1.1.5), also known as lysophospholipase, lecithinase B, or lysolecithinase are widely distributed enzymes that metabolize intracellular lipids, and occur in numerous isoforms. Small isoforms, approximately 15-30 kD, function as hydrolases; large isoforms, those exceeding 60 kD, function both as hydrolases and transacylases. A particular

substrate for PLBs, lysophosphatidylcholine, causes lysis of cell membranes when it is formed or imported into a cell. PLBs are regulated by lipid factors including acylcarnitine, arachidonic acid, and phosphatidic acid. These lipid factors are signaling molecules important in numerous pathways, including the inflammatory response (Anderson, R. et al. (1994) *Toxicol. Appl. Pharmacol.* 125:176-183; Selle, H. et al. (1993); *Eur. J. Biochem.* 212:411-416).

Phospholipase C (PLC) (ExPASy ENZYME EC 3.1.4.10) plays an important role in transmembrane signal transduction. Many extracellular signaling molecules including hormones, growth factors, neurotransmitters, and immunoglobulins bind to their respective cell surface receptors and activate PLCs. The role of an activated PLC is to catalyze the hydrolysis of phosphatidyl-inositol-4, 5-bisphosphate (PIP<sub>2</sub>), a minor component of the plasma membrane, to produce diacylglycerol and inositol 1, 4, 5-trisphosphate (IP<sub>3</sub>). In their respective biochemical pathways, IP<sub>3</sub> and diacylglycerol serve as second messengers and trigger a series of intracellular responses. IP<sub>3</sub> induces the release of Ca<sup>2+</sup> from internal cellular storage, and diacylglycerol activates protein kinase C (PKC). Both pathways are part of transmembrane signal transduction mechanisms which regulate cellular processes which include secretion, neural activity, metabolism, and proliferation.

Several distinct isoforms of PLC have been identified and are categorized as PLC-beta, PLC-gamma, and PLC-delta. Subtypes are designated by adding Arabic numbers after the Greek letters, eg. PLC- $\beta$ -1. PLCs have a molecular mass of 62-68 kDa, and their amino acid sequences show two regions of significant similarity. The first region designated X has about 170 amino acids, and the second or Y region contains about 260 amino acids.

The catalytic activities of the three isoforms of PLC are dependent upon Ca<sup>2+</sup>. It has been suggested that the binding sites for Ca<sup>2+</sup> in the PLCs are located in the Y-region, one of two conserved regions. The hydrolysis of common inositol-containing phospholipids, such as phosphatidylinositol (PI), phosphatidylinositol 4-monophosphate (PIP), and phosphatidylinositol 4, 5-bisphosphate (PIP<sub>2</sub>), by any of the isoforms yields cyclic and noncyclic inositol phosphates (Rhee, S.G. and Y.S. Bae (1997) *J. Biol. Chem.* 272:15045-15048).

All mammalian PLCs contain a pleckstrin homology (PH) domain which is about 100 amino acids in length and is composed of two antiparallel beta sheets flanked by an amphipathic alpha helix. PH domains target PLCs to the membrane surface by interacting with either the beta/gamma subunits of G proteins or PIP<sub>2</sub> (PROSITE PDOC50003).

Phospholipase D (PLD) (ExPASy ENZYME EC 3.1.4.4), also known as lecithinase D, lipophosphodiesterase II, and choline phosphatase catalyzes the hydrolysis of phosphatidylcholine and



other phospholipids to generate phosphatidic acid. PLD plays an important role in membrane vesicle trafficking, cytoskeletal dynamics, and transmembrane signal transduction. In addition, the activation of PLD is involved in cell differentiation and growth (reviewed in Liscovitch, M. (2000) *Biochem. J.* 345:401-415).

- 5 PLD is activated in mammalian cells in response to diverse stimuli that include hormones, neurotransmitters, growth factors, cytokines, activators of protein kinase C, and agonists binding to G-protein-coupled receptors. At least two forms of mammalian PLD, PLD1 and PLD2, have been identified. PLD1 is activated by protein kinase C alpha and by the small GTPases ARF and RhoA. (Houle, M.G. and S. Bourgoin (1999) *Biochim. Biophys. Acta* 1439:135-149). PLD2 can be  
10 selectively activated by unsaturated fatty acids such as oleate (Kim, J.H. (1999) *FEBS Lett.* 454:42-46).

#### Lipoxygenases

Lipoxygenases (ExPASy ENZYME EC 1.13.11.12) are non-heme iron-containing enzymes that catalyze the dioxygenation of certain polyunsaturated fatty acids such as lipoproteins.

- 15 Lipoxygenases are found widely in plants, fungi, and animals. Several different lipoxygenase enzymes are known, each having a characteristic oxidation action. In animals, there are specific lipoxygenases that catalyze the dioxygenation of arachidonic acid at the carbon-3, 5, 8, 11, 12, and 15 positions. These enzymes are named after the position of arachidonic acid that they dioxygenate. Lipoxygenases have a single polypeptide chain with a molecular mass of ~75-80 kDa in animals. The  
20 proteins have an N-terminal-barrel domain and a larger catalytic domain containing a single atom of non-heme iron. Oxidation of the ferric enzyme to an active form is required for catalysis (Yamamoto, S. (1992) *Biochim. Biophys. Acta* 1128:117-131; Brash, A.R. (1999) *J. Biol. Chem.* 274:23679-23682). A variety of lipoxygenase inhibitors exist and are classified into five major categories according to their mechanism of inhibition. These include antioxidants, iron chelators,  
25 substrate analogues, lipoxygenase-activating protein inhibitors, and, finally, epidermal growth factor-receptor inhibitors.

3-Lipoxygenase, also known as e-LOX-3 or Aloxe3 has recently been cloned from murine epidermis. Aloxe3 resides on mouse chromosome 11, and the deduced amino acid sequence for Aloxe3 is 54% identical to the 12-lipoxygenase sequences (Kinzig, A. (1999) *Genomics* 58:158-164).

- 30 5-Lipoxygenase (5-LOX, ExPASy ENZYME EC 1.13.11.34), also known as arachidonate:oxygen 5-oxidoreductase, is found primarily in white blood cells, macrophages, and mast cells. 5-LOX converts arachidonic acid first to 5-hydroperoxyeicosatetraenoic acid (5-HPETE) and then to leukotriene (LTA4 (5,6-oxido-7,9,11,14-eicosatetraenoic acid)). Subsequent conversion of

leukotriene A4 by leukotriene A4 hydrolase yields the potent neutrophil chemoattractant leukotriene B4. Alternatively, conjugation of LTA4 with glutathione by leukotriene C4 synthase plus downstream metabolism leads to the cysteinyl leukotrienes that influence airway reactivity and mucus secretion, especially in asthmatics. Most lipoxygenases require no other cofactors or proteins for activity. In contrast, the mammalian 5-LOX requires calcium and ATP, and is activated in the presence of a 5-LOX activating protein (FLAP). FLAP itself binds to arachidonic acid and supplies 5-LOX with substrate (Lewis, R.A. et al. (1990) *New Engl. J. Med.* 323:645-655). The expression levels of 5-LOX and FLAP are found to be increased in the lungs of patients with plexogenic (primary) pulmonary hypertension (Wright, L. et al. (1998) *Am. J. Respir. Crit. Care Med.* 157:219-229).

12-Lipoxygenase (12-LOX, ExPASy ENZYME: EC 1.13.11.31) oxygenates arachidonic acid to form 12-hydroperoxyeicosatetraenoic acid (12-HPETE). Mammalian 12-lipoxygenases are named after the prototypical tissues of their occurrence (hence, the leukocyte, platelet, or epidermal types). Platelet-type 12-LOX has been found to be the predominant isoform in epidermal skin specimens and epidermoid cells. Leukocyte 12-LOX was first characterized extensively from porcine leukocytes and was found to have a rather broad distribution in mammalian tissues by immunochemical assays. Besides tissue distribution, the leukocyte 12-LOX is distinguished from the platelet-type enzyme by its ability to form 15-HPETE, in addition to 12-HPETE from arachidonic acid substrate. Leukocyte 12-LOX is highly related to 15-lipoxygenase (15-LOX) in that both are dual specificity lipoxygenases, and they are about 85% identical in primary structure in higher mammals. Leukocyte 12-LOX is found in tracheal epithelium, leukocytes, and macrophages (Conrad, D.J. (1999) *Clin. Rev. Allergy Immunol.* 17:71-89).

15-Lipoxygenase (15-LOX; ExPASy ENZYME: EC 1.13.11.33) is found in human reticulocytes, airway epithelium, and eosinophils. 15-LOX has been detected in atherosclerotic lesions in mammals, specifically rabbit and man. The enzyme, in addition to its role in oxidative modification of lipoproteins, is important in the inflammatory reaction in atherosclerotic lesions. 15-LOX has been shown to be induced in human monocytes by the cytokine IL-4, which is known to be implicated in the inflammatory process (Kuhn, H. and S. Borngraber (1999) *Adv. Exp. Med. Biol.* 447:5-28).

#### Disease Correlation

Lipid metabolism is involved in human diseases and disorders. In the arterial disease atherosclerosis, fatty lesions form on the inside of the arterial wall. These lesions promote the loss of arterial flexibility and the formation of blood clots (Guyton, *supra*). In Tay-Sachs disease, the GM<sub>2</sub> ganglioside (a sphingolipid) accumulates in lysosomes of the central nervous system due to a lack of the enzyme N-acetylhexosaminidase. Patients suffer nervous system degeneration leading to early

death (Fauci, A.S. et al. (1998) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, p. 2171). The Niemann-Pick diseases are caused by defects in lipid metabolism. Niemann-Pick diseases types A and B are caused by accumulation of sphingomyelin (a sphingolipid) and other lipids in the central nervous system due to a defect in the enzyme sphingomyelinase, leading to

- 5 neurodegeneration and lung disease. Niemann-Pick disease type C results from a defect in cholesterol transport, leading to the accumulation of sphingomyelin and cholesterol in lysosomes and a secondary reduction in sphingomyelinase activity. Neurological symptoms such as grand mal seizures, ataxia, and loss of previously learned speech, manifest 1-2 years after birth. A mutation in the NPC protein, which contains a putative cholesterol-sensing domain, was found in a mouse model of
- 10 Niemann-Pick disease type C (Fauci, supra, p. 2175; Loftus, S.K. et al. (1997) *Science* 277:232-235).

PLAs are implicated in a variety of disease processes. For example, PLAs are found in the pancreas, in cardiac tissue, and in inflammation-associated tissues. Pancreatic PLAs function in the digestion of dietary lipids and have been proposed to play a role in cell proliferation, smooth muscle contraction, and acute lung injury. Inflammatory PLAs are potent mediators of inflammatory

15 processes and are highly expressed in serum and synovial fluids of patients with inflammatory disorders. Additionally, inflammatory PLAs are found in most human cell types and are expressed in diverse pathological processes such as septic shock, intestinal cancers, rheumatoid arthritis, and epidermal hyperplasia.

The role of PLBs in human tissues has been investigated in various research studies.

20 Hydrolysis of lysophosphatidylcholine by PLBs causes lysis in erythrocyte membranes (Selle, supra). Similarly, Endresen, M.J. et al. (1993; *Scand. J. Clin. Invest.* 53:733-739) reported that the increased hydrolysis of lysophosphatidylcholine by PLB in pre-eclamptic women causes release of free fatty acids into the sera. In renal studies, PLB was shown to protect Na<sup>+</sup>,K<sup>+</sup>-ATPase from the cytotoxic and cytolytic effects of cyclosporin A (Anderson, supra).

25 Lipases, phospholipases, and lipoxygenases are thought to contribute to complex diseases, such as atherosclerosis, obesity, arthritis, asthma, and cancer, as well as to single gene defects, such as Wolman's disease and Type I hyperlipoproteinemia.

The discovery of new lipid metabolism enzymes, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention,

30 and treatment of cancer, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, and cardiovascular disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of lipid metabolism enzymes.

## SUMMARY OF THE INVENTION

The invention features purified polypeptides, lipid metabolism enzymes, referred to collectively as "LMM" and individually as "LMM-1," "LMM-2," "LMM-3," "LMM-4," "LMM-5," and "LMM-6." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of

5 a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence

10 selected from the group consisting of SEQ ID NO:1-6. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-6.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence

15 at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-6. In

20 another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:7-12.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least

25 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention

30 provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least

90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) culturing.  
5 a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence  
10 selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ  
15 ID NO:1-6.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, c)  
20 a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of  
25 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the  
30 sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of

said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional LMM, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a

pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional LMM, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as  
5 an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and  
10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of  
15 treating a disease or condition associated with overexpression of functional LMM, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a  
20 naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6. The method comprises a) combining the polypeptide with at least one  
25 test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, b) a polypeptide comprising a  
30 naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group

consisting of SEQ ID NO:1-6. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.



### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now

described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## 5 DEFINITIONS

"LMM" refers to the amino acid sequences of substantially purified LMM obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of  
10 LMM. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of LMM either by directly interacting with LMM or by acting on components of the biological pathway in which LMM participates.

An "allelic variant" is an alternative form of the gene encoding LMM. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in  
15 polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

20 "Altered" nucleic acid sequences encoding LMM include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as LMM or a polypeptide with at least one functional characteristic of LMM. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding LMM, and improper or unexpected hybridization to allelic variants, with a  
25 locus other than the normal chromosomal locus for the polynucleotide sequence encoding LMM. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent LMM. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological  
30 or immunological activity of LMM is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains

having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic  
5 molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known  
10 in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of LMM. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of LMM either by directly interacting with LMM or by acting on components of the biological pathway in which LMM  
15 participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind LMM polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used  
20 to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that  
25 makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

30 The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified

sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring  
5 nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic"  
10 refers to the capability of the natural, recombinant, or synthetic LMM, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement,  
15 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding LMM or fragments of LMM may be  
20 employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to  
25 repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended  
30 and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino

acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
5	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
10	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
15	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
20	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a

diseased and a normal sample.

“Exon shuffling” refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A “fragment” is a unique portion of LMM or the polynucleotide encoding LMM which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:7-12 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:7-12, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:7-12 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:7-12 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:7-12 and the region of SEQ ID NO:7-12 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-6 is encoded by a fragment of SEQ ID NO:7-12. A fragment of SEQ ID NO:1-6 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-6. For example, a fragment of SEQ ID NO:1-6 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-6. The precise length of a fragment of SEQ ID NO:1-6 and the region of SEQ ID NO:1-6 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A “full length” polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A “full

length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to  
5 the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default  
10 parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as  
15 follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment  
20 Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2  
25 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version  
30 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

5 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous  
10 nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes  
15 in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment  
20 methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e  
25 sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

30 Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:



*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

5 *Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for  
10 instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain  
15 DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

20 "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the  
25 stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity.  
30 Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about

5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989)

5 Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may  
10 be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily  
15 apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A  
20 hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide  
25 sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

30 An "immunogenic fragment" is a polypeptide or oligopeptide fragment of LMM which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of LMM which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

5       The term "modulate" refers to a change in the activity of LMM. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of LMM.

10       The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

15       "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

20       "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

25       "Post-translational modification" of an LMM may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of LMM.

30       "Probe" refers to nucleic acid sequences encoding LMM, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers  
5 may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular  
10 Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

15 Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU  
20 primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to  
25 avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing  
30 selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example,

as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence  
5 that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a  
10 recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

15 A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid,  
20 amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of  
25 the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing LMM, nucleic acids encoding LMM, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in  
30 solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure

of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

5           The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

          A "substitution" refers to the replacement of one or more amino acid residues or nucleotides  
10 by different amino acid residues or nucleotides, respectively.

          "Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

15           A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

          "Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid  
20 sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently  
25 transformed cells which express the inserted DNA or RNA for limited periods of time.

          A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell,  
30 by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria,

fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

5 A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least  
10 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding  
15 polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide  
20 polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of  
25 the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence  
30 identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human lipid metabolism enzymes (LMM), the polynucleotides encoding LMM, and the use of these compositions for the diagnosis, treatment, or

prevention of cancer, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, and cardiovascular disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are lipid metabolism enzymes. For example, SEQ ID NO:1 is 53% identical to human lysosomal acid lipase (GenBank ID g505053) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.2e-116, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains an alpha/beta hydrolase fold as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of



conserved protein domains. (See Table 3.) Data from MOTIFS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is a lipase. In an alternative example, SEQ ID NO:4 is 63% identical to human pregastric lipase (GenBank ID g579561) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.4e-142, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:4 also contains an alpha/beta hydrolase fold as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from MOTIFS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:4 is a lipase. In an alternative example, SEQ ID NO:5 is 35% identical to human arylacetamide deacetylase (GenBank ID g537514) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.6e-47, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. In a further example, SEQ ID NO:6 is 51% identical to murine phospholipase C-L2 (GenBank ID g6705987) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.1e-151, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains phosphatidylinositol-specific phospholipase X and Y domains, a C2 domain, and a pleckstrin homology domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:6 is a phospholipase. SEQ ID NO:2 and SEQ ID NO:3 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-6 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:7-12 or that distinguish between SEQ ID NO:7-12 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length

polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 6875328H1 is the identification number of an Incyte cDNA sequence, and EPIMUNN04 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71075936V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g5369512) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>\_YYYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3,...</sub> if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FLXXXXXX\_gAAAAA\_gBBBBB\_1\_N is the identification number of a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis

methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses LMM variants. A preferred LMM variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the LMM amino acid sequence, and which contains at least one functional or structural characteristic of LMM.

The invention also encompasses polynucleotides which encode LMM. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:7-12, which encodes LMM. The polynucleotide sequences of SEQ ID NO:7-12, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding LMM. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding LMM. A particular aspect of the invention encompasses a variant of a

polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:7-12 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:7-12. Any one of the polynucleotide variants described above can encode an amino acid  
5 sequence which contains at least one functional or structural characteristic of LMM.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding LMM, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be  
10 made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring LMM, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode LMM and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring LMM under appropriately selected  
15 conditions of stringency, it may be advantageous to produce nucleotide sequences encoding LMM or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide  
20 sequence encoding LMM and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode LMM and LMM derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the  
25 synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding LMM or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID  
30 NO:7-12 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of

the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding LMM may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about

68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode LMM may be cloned in recombinant DNA molecules that direct expression of LMM, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express LMM.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter LMM-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of LMM, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is

produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

10 In another embodiment, sequences encoding LMM may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser. 7*:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser. 7*:225-232.) Alternatively, LMM itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g.,  
15 Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of LMM, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a  
20 sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

25 In order to express a biologically active LMM, the nucleotide sequences encoding LMM or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences  
30 encoding LMM. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding LMM. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding LMM and its initiation codon and upstream regulatory sequences are inserted into

the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both  
5 natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding LMM and appropriate transcriptional and translational control  
10 elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences  
15 encoding LMM. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or  
20 animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and  
25 Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994)  
30 *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding LMM. For example, routine cloning,



subcloning, and propagation of polynucleotide sequences encoding LMM can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPO1 plasmid (Life Technologies). Ligation of sequences encoding LMM into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of

5 transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of LMM are needed, e.g. for the production of antibodies, vectors which direct high level expression of LMM may be used. For example, vectors  
10 containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of LMM. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration  
15 of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of LMM. Transcription of sequences encoding LMM may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used  
20 alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated  
25 transfection. (See, e.g., *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding LMM may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader  
30 sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses LMM in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-

based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of LMM in cell lines is preferred. For example, sequences encoding LMM can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *ap<sup>r</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding LMM is inserted within a marker gene sequence, transformed cells containing

sequences encoding LMM can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding LMM under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

5 In general, host cells that contain the nucleic acid sequence encoding LMM and that express LMM may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

10 Immunological methods for detecting and measuring the expression of LMM using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on LMM is preferred, but a competitive  
15 binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

20 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding LMM include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding LMM, or any fragments thereof, may be cloned into a vector  
25 for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for  
30 ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding LMM may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein

produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode LMM may be designed to contain signal sequences which direct secretion of LMM through a prokaryotic or eukaryotic cell membrane.

5 In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells  
10 which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid  
15 sequences encoding LMM may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric LMM protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of LMM activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity  
20 matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion  
25 proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the LMM encoding sequence and the heterologous protein sequence, so that LMM may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially  
30 available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled LMM may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the

T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

LMM of the present invention or fragments thereof may be used to screen for compounds that specifically bind to LMM. At least one and up to a plurality of test compounds may be screened  
5 for specific binding to LMM. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of LMM, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):  
10 Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which LMM binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express LMM, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E.  
15 coli. Cells expressing LMM or cell membrane fractions which contain LMM are then contacted with a test compound and binding, stimulation, or inhibition of activity of either LMM or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the  
20 assay may comprise the steps of combining at least one test compound with LMM, either in solution or affixed to a solid support, and detecting the binding of LMM to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

LMM of the present invention or fragments thereof may be used to screen for compounds  
25 that modulate the activity of LMM. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for LMM activity, wherein LMM is combined with at least one test compound, and the activity of LMM in the presence of a test compound is compared with the activity of LMM in the absence of the test  
30 compound. A change in the activity of LMM in the presence of the test compound is indicative of a compound that modulates the activity of LMM. Alternatively, a test compound is combined with an in vitro or cell-free system comprising LMM under conditions suitable for LMM activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of LMM may do

so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding LMM or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding LMM may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) *Science* 282:1145-1147).

Polynucleotides encoding LMM can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding LMM is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress LMM, e.g., by secreting LMM in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between

regions of LMM and lipid metabolism enzymes. In addition, the expression of LMM is closely associated with leukocytes, pituitary glands, and reproductive and bronchial tissue. In the treatment of disorders associated with increased LMM expression or activity, it is desirable to decrease the expression or activity of LMM. In the treatment of disorders associated with decreased LMM

5 expression or activity, it is desirable to increase the expression or activity of LMM.

Therefore, in one embodiment, LMM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LMM.

Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of  
10 the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders,  
15 amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial  
20 insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders,  
25 dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an  
30 autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact

dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or

5 pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a gastrointestinal disorder such as

10 dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma,

15 infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease,  $\alpha_1$ -antitrypsin deficiency, Reye's syndrome, primary

20 sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a skin disorder, such as dermatitis, eczema, ichthyosis, keratosis, psoriasis, scleroderma, and skin atrophy; a disorder of lipid metabolism such as fatty liver,

25 cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis,

30 disseminated fat necrosis, adiposis dolorosa, lipid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-



Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation.

In another embodiment, a vector capable of expressing LMM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LMM including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified LMM in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LMM including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of LMM may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LMM including, but not limited to, those listed above.

In a further embodiment, an antagonist of LMM may be administered to a subject to treat or

prevent a disorder associated with increased expression or activity of LMM. Examples of such disorders include, but are not limited to, those cancer, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, and cardiovascular disorders described above. In one aspect, an antibody which specifically binds LMM may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express LMM.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding LMM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of LMM including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of LMM may be produced using methods which are generally known in the art. In particular, purified LMM may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind LMM. Antibodies to LMM may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with LMM or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to LMM have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are

identical to a portion of the amino acid sequence of the natural protein. Short stretches of LMM amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to LMM may be prepared using any technique which provides for the  
5 production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

10 In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single  
15 chain antibodies may be adapted, using methods known in the art, to produce LMM-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population  
20 or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for LMM may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin  
25 digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired  
30 specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between LMM and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two

non-interfering LMM epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for LMM. Affinity is expressed as an association  
5 constant,  $K_a$ , which is defined as the molar concentration of LMM-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple LMM epitopes, represents the average affinity, or avidity, of the antibodies for LMM. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular LMM  
10 epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the LMM-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of LMM, preferably in active form, from the antibody (Catty, D.  
15 (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg  
20 specific antibody/ml, is generally employed in procedures requiring precipitation of LMM-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding LMM, or any fragment  
25 or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding LMM. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding  
30 LMM. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence

complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 5 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

10 In another embodiment of the invention, polynucleotides encoding LMM may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency 15 (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) 20 express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA.* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides* 25 *brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in LMM expression or regulation causes disease, the expression of LMM from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in 30 LMM are treated by constructing mammalian expression vectors encoding LMM and introducing these vectors by mechanical means into LMM-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and

(v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of LMM include, but are not  
5 limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors  
(Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA),  
and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). LMM  
may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous  
sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter  
10 (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci.*  
*USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau  
(1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen));  
the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the  
FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V.  
15 and Blau, H.M. *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous  
gene encoding LMM from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID  
TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver  
polynucleotides to target cells in culture and require minimal effort to optimize experimental  
20 parameters. In the alternative, transformation is performed using the calcium phosphate method  
(Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al.  
(1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these  
standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with  
25 respect to LMM expression are treated by constructing a retrovirus vector consisting of (i) the  
polynucleotide encoding LMM under the control of an independent promoter or the retrovirus long  
terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive  
element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences  
required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are  
30 commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc.*  
*Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in  
an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for  
receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al.

(1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses  
5 a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998)  
10 Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding LMM to cells which have one or more genetic abnormalities with respect to the expression of LMM. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be  
15 versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 389:239-242, both  
20 incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding LMM to target cells which have one or more genetic abnormalities with respect to the expression of LMM. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing LMM to cells of the central nervous system, for which HSV has a  
25 tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is  
30 hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4,

ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and  
5 propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding LMM to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based  
10 on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for LMM into the alphavirus  
15 genome in place of the capsid-coding region results in the production of a large number of LMM-coding RNAs and the synthesis of high levels of LMM in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application  
20 (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of LMM into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

25 Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have  
30 been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.



Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze  
5 endonucleolytic cleavage of sequences encoding LMM.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for  
10 secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques  
15 for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding LMM. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells,  
20 or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be  
25 extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding LMM. Compounds  
30 which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective

compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased LMM expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding LMM may be therapeutically useful, and in the treatment of disorders associated with  
5 decreased LMM expression or activity, a compound which specifically promotes expression of the polynucleotide encoding LMM may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in  
10 altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding LMM is exposed to at least one test compound thus obtained. The sample  
15 may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding LMM are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding LMM. The amount of hybridization may be quantified, thus forming the basis  
20 for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al.  
25 (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al.  
30 (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient.

Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of  
5 such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various  
10 formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of LMM, antibodies to LMM, and mimetics, agonists, antagonists, or inhibitors of LMM.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal,  
15 intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-  
20 acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

25 Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising LMM or fragments thereof. For example, liposome preparations  
30 containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, LMM or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et

al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and  
5 route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example LMM or fragments thereof, antibodies of LMM, and agonists, antagonists or inhibitors of LMM, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by  
10 standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are  
15 used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the  
20 subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or  
25 biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their  
30 inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind LMM may be used for the

diagnosis of disorders characterized by expression of LMM, or in assays to monitor patients being treated with LMM or agonists, antagonists, or inhibitors of LMM. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for LMM include methods which utilize the antibody and a label to detect LMM in human body  
5 fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring LMM, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of LMM expression. Normal  
10 or standard values for LMM expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to LMM under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of LMM expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation  
15 between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding LMM may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of LMM may be correlated with  
20 disease. The diagnostic assay may be used to determine absence, presence, and excess expression of LMM, and to monitor regulation of LMM levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding LMM or closely related molecules may be used to identify nucleic acid sequences which encode LMM. The specificity of the probe, whether it is made  
25 from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding LMM, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50%  
30 sequence identity to any of the LMM encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:7-12 or from genomic sequences including promoters, enhancers, and introns of the LMM gene.

Means for producing specific hybridization probes for DNAs encoding LMM include the

cloning of polynucleotide sequences encoding LMM or LMM derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a  
5 variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding LMM may be used for the diagnosis of disorders associated with expression of LMM. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,  
10 teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other  
15 extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker  
20 syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular  
25 disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial  
30 frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis,

cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis,

5 myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral,

10 bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma,

15 infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary

20 sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a skin disorder, such as dermatitis, eczema, ichthyosis, keratosis, psoriasis, scleroderma, and skin atrophy; a disorder of lipid metabolism such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency,

25 myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis,

30 disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-

Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation. The polynucleotide sequences encoding LMM may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered LMM expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding LMM may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding LMM may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding LMM in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the



efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of LMM, a normal or standard profile for expression is established. This may be accomplished by combining  
5 body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding LMM, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from  
10 patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from  
15 successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual  
20 clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding LMM may involve the use of PCR. These oligomers may be chemically synthesized, generated  
25 enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding LMM, or a fragment of a polynucleotide complementary to the polynucleotide encoding LMM, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

30 In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding LMM may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation

polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding LMM are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the

5 secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual  
10 overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

15 Methods which may also be used to quantify the expression of LMM include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of  
20 interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large  
25 numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used  
30 to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, LMM, fragments of LMM, or antibodies specific for LMM may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to  
5 generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by  
10 hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

15 Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of  
20 pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a  
25 signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the  
30 rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for

example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

5           In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an  
10           untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

          Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome  
15           can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by  
20           isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently  
25           positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial  
30           sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

          A proteomic profile may also be generated using antibodies specific for LMM to quantify the

levels of LMM expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoze, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a  
5 variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor  
10 correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

15 In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the  
20 test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated  
25 with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g.,  
30 Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of

microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding LMM may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either  
5 coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs),  
10 yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the  
15 inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man  
20 (OMIM) World Wide Web site. Correlation between the location of the gene encoding LMM on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps.  
25 Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any  
30 sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, LMM, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes  
5 between LMM and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with LMM, or fragments thereof,  
10 and washed. Bound LMM is then detected by methods well known in the art. Purified LMM can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing  
15 antibodies capable of binding LMM specifically compete with a test compound for binding LMM. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with LMM.

In additional embodiments, the nucleotide sequences which encode LMM may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on  
20 properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way  
25 whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/227,429, U.S. Ser. No. 60/231,370, U.S. Ser. No. 60/233,212, and U.S. Ser. No. 60/236,885, are hereby expressly incorporated by reference.

## 30 EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were

homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol  
5 or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively,  
10 RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIP<sup>T</sup> plasmid system (Life Technologies), using  
15 the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column  
20 chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA (Invitrogen), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were  
25 transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using  
30 at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1



ml of distilled water and stored, with or without lyophilization, at 4 °C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 5 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

### III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. 10 Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as 15 the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading 20 frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and 25 programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary 30 structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or

Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive  
5 the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide  
10 sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

15 Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where  
20 applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID  
25 NO:7-12. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative lipid metabolism enzymes were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpr1 and gbhtg). Genscan is  
30 a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of

Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode lipid metabolism enzymes, the encoded polypeptides were analyzed by querying against PFAM models for lipid metabolism enzymes. Potential lipid metabolism enzymes were also identified by homology to Incyte cDNA sequences that had been annotated as lipid metabolism enzymes. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpr public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

#### V. Assembly of Genomic Sequence Data with cDNA Sequence Data

##### "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent

type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

5 **"Stretched" Sequences**

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST  
10 analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous  
15 genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

**VI. Chromosomal Mapping of LMM Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:7-12 were compared with  
20 sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:7-12 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for  
25 Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-  
30 arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid

markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

## 5 VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

10 Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

15

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the  
 20 length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by  
 25 gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced  
 30 either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding LMM are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are

assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding LMM. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

#### VIII. Extension of LMM Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4

repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, 5 Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

10 The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended 15 clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

20 The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA 25 recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or 30 are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:7-12 are employed to screen cDNAs,

genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of  
5 [  $\gamma$ - $^{32}$ P ] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase  
10 I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate.  
15 Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

#### X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra.*), mechanical  
20 microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Skena (1999), *supra.*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may  
25 be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Skena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may  
30 comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection.



After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C

oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

### Hybridization

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source,

although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

#### **XI. Complementary Polynucleotides**

Sequences complementary to the LMM-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring LMM. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of LMM. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the LMM-encoding transcript.

#### **XII. Expression of LMM**

Expression and purification of LMM is achieved using bacterial or virus-based expression systems. For expression of LMM in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA

transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).

Antibiotic resistant bacteria express LMM upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of LMM in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding LMM by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, LMM is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from LMM at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified LMM obtained by these methods can be used directly in the assays shown in Examples XVI and XVII where applicable.

### XIII. Functional Assays

LMM function is assessed by expressing the sequences encoding LMM at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a

marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of LMM on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding LMM and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding LMM and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### XIV. Production of LMM Specific Antibodies

LMM substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the LMM amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-

Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-LMM activity by, for example, binding the peptide or LMM to a substrate,  
5 blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### XV. Purification of Naturally Occurring LMM Using Specific Antibodies

Naturally occurring or recombinant LMM is substantially purified by immunoaffinity chromatography using antibodies specific for LMM. An immunoaffinity column is constructed by  
10 covalently coupling anti-LMM antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing LMM are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of LMM (e.g., high ionic strength  
15 buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/LMM binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and LMM is collected.

#### XVI. Identification of Molecules Which Interact with LMM

LMM, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent.  
20 (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled LMM, washed, and any wells with labeled LMM complex are assayed. Data obtained using different concentrations of LMM are used to calculate values for the number, affinity, and association of LMM with the candidate molecules.

25 Alternatively, molecules interacting with LMM are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

LMM may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions  
30 between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

#### XVII. Demonstration of LME Activity

LME activity can be demonstrated by an in vitro hydrolysis assay with vesicles containing

1-palmitoyl-2-[1-<sup>14</sup>C]oleoyl phosphatidylcholine (Sigma-Aldrich). LME triglyceride lipase activity and phospholipase A<sub>2</sub> activity are demonstrated by analysis of the cleavage products isolated from the hydrolysis reaction mixture.

Vesicles containing 1-palmitoyl-2-[1-<sup>14</sup>C]oleoyl phosphatidylcholine (Amersham Pharmacia Biotech.) are prepared by mixing 2.0 μCi of the radiolabeled phospholipid with 12.5 mg of unlabeled 1-palmitoyl-2-oleoyl phosphatidylcholine and drying the mixture under N<sub>2</sub>. 2.5 ml of 150 mM Tris-HCl, pH 7.5, is added, and the mixture is sonicated and centrifuged. The supernatant may be stored at 4 °C. The final reaction mixtures contain 0.25 ml of Hanks buffered salt solution supplemented with 2.0 mM taurochenodeoxycholate, 1.0% bovine serum albumin, 1.0 mM CaCl<sub>2</sub>, pH 7.4, 150 μg of 1-palmitoyl-2-[1-<sup>14</sup>C]oleoyl phosphatidylcholine vesicles, and various amounts of LME diluted in PBS. After incubation for 30 min at 37 °C, 20 μg each of lyso-phosphatidylcholine and oleic acid are added as carriers and each sample is extracted for total lipids. The lipids are separated by thin layer chromatography using a two solvent system of chloroform:methanol:acetic acid:water (65:35:8:4) until the solvent front is halfway up the plate. The process is then continued with hexane:ether:acetic acid (86:16:1) until the solvent front is at the top of the plate. The lipid-containing areas are visualized with I<sub>2</sub> vapor; the spots are scraped, and their radioactivity is determined by scintillation counting. The amount of radioactivity released as fatty acids will increase as a greater amount of LME is added to the assay mixture while the amount of radioactivity released as lyso-phosphatidylcholine will remain low. This demonstrates that LME cleaves at the sn-2 and not the sn-1 position, as is characteristic of phospholipase A<sub>2</sub> activity.

Alternatively, LME phospholipase activity is measured by the hydrolysis of a fatty acyl residue at the sn-1 position of phosphatidylserine. LME is combined with the Tritium [<sup>3</sup>H] labeled substrate phosphatidylserine at stoichiometric quantities in a suitable buffer. Following an appropriate incubation time, the hydrolyzed reaction products are separated from the substrates by chromatographic methods. The amount of acylglycerophosphoserine produced is measured by counting tritiated product with the help of a scintillation counter. Various control groups are set up to account for background noise and unincorporated substrate. The final counts represent the tritiated enzyme product [<sup>3</sup>H]-acylglycerophosphoserine, which is directly proportional to the activity of LME in biological samples.

LME lipoxygenase activity can be measured by chromatographic methods. Extracted LME lipoxygenase protein is incubated with 100 μM [1-<sup>14</sup>C] arachidonic acid or other unlabeled fatty acids at 37°C for 30 min. After the incubation, stop solution (acetonitrile:methanol:water, 350:150:1) is added. The samples are extracted and analyzed by reverse-phase HPLC by using a solvent system of methanol/water/acetic acid, 85:15:0.01 (vol/vol) at a flow rate of 1 ml/min. The effluent is monitored

at 235 nm and analyzed for the presence of the major arachidonic metabolite such as 12-HPETE (catalyzed by 12-LOX). The fractions are also subjected to liquid scintillation counting. The final counts represent the products, which is directly proportional to the activity of LME in biological samples. For stereochemical analysis, the metabolites of arachidonic acid are analyzed further by  
5 chiral phase-HPLC and by mass spectrometry (Sun, D. et al. (1998) J. Biol. Chem. 273:33540-33547).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments.  
10 Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.



Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7479063	1	7479063CD1	7	7479063CB1
7479064	2	7479064CD1	8	7479064CB1
7478088	3	7478088CD1	9	7478088CB1
7480935	4	7480935CD1	10	7480935CB1
7481764	5	7481764CD1	11	7481764CB1
72023748	6	72023748CD1	12	72023748CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	7479063CD1	g505053	5.20E-116	[Homo sapiens] lysosomal acid lipase
2	7479064CD1	g14749	2.30E-97	[Rattus norvegicus] prelingual lipase
3	7478088CD1	g219393	4.10E-08	arachidonate 12-lipoxygenase [Homo sapiens] (Yoshimoto, T. et al. (1990) Biochem. Biophys. Res. Commun. 172: 1230-1235; Yoshimoto, T. et al. (1992) J. Biol. Chem. 267: 24805-24809)
4	7480935CD1	g579561	1.40E-142	pregastric lipase [Homo sapiens] (Tubb, et al. (1986) Patent Number: WO 8603778-A)
5	7481764CD1	g537514	1.60E-47	arylacetamide deacetylase [Homo sapiens] (Probst, M.R. et al. (1994) J. Biol. Chem. 269: 21650- 21656)
6	72023748CD1	g6705987	2.10E-151	phospholipase C-L2 [Mus musculus] (Otsuki, M. et al. (1999) Biochem. Biophys. Res. Commun. 266: 97-103)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7479063CD1	403	S130 S137 S147 S209 S248 S271 S279 T135 T167 T190 T353 T58 Y173	N105 N165 N277 N325 N40	Lipases, serine active site: A152-N201 alpha/beta hydrolase fold abhydrolase: Y117-I396 TRIACYLGLYCEROL LIPASE, LINGUAL DM02342 P38571 3-397: M7-Y402 TRIACYLGLYCEROL LIPASE, LINGUAL DM02342 P07098 35-395: S42-M399 TRIACYLGLYCEROL LIPASE, LINGUAL DM02342 JC4017 1-394: M7-M399 TRIACYLGLYCEROL LIPASE, LINGUAL DM02342 P04634 32-394: M39-M399 LIPASE HYDROLASE PRECURSOR SIGNAL LIPID DEGRADATION PROTEIN GLYCOPROTEIN ESTERASE TRIACYLGLYCEROL PD003556: F26-M399	PROFILES SCAN HMMER_PFAM BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_PRODOM
2	7479064CD1	289	S109 S124 S131 S141 S218 S263 T184 T203 T215 T235 T52	N100 N34 N99	Lipases, serine active site: L172-T181 SpSCAN signal cleavage: M1-N22 Lipases, serine active site: A146-F195 TRIACYLGLYCEROL LIPASE, LINGUAL DM02342 P04634 32-394: M33-M272 JC4017 1-394: M1-N282 P07098 35-395: I35-N271 P38571 3-397: N28-M272 LIPASE HYDROLASE PRECURSOR SIGNAL LIPID DEGRADATION PROTEIN GLYCOPROTEIN ESTERASE TRIACYLGLYCEROL PD003556: N28-N271 signal peptide: M1-G19	MOTIFS SPSCAN PROFILES SCAN BLAST_DOMO BLAST_PRODOM
3	7478088CD1	168	S117 S157 S6 T135 T60		Lipases, serine active site: L166-T175 signal cleavage: M1-G19 Signal peptide: M1-A39 PLAT (Polycystin-1, Lipoxigenase, Alpha-Toxin)/LH2 (Lipoxigenase homology) domain: V44-E160	HMMER MOTIFS SPSCAN SPSCAN HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	7480935CD1	399	S109 S124 S131 S141 S218 S263 S273 T184 T203 T215 T235 T329 T52	N271 N327 N34 N99	signal peptide: M1-G19 Signal peptide: M1-G19 alpha/beta hydrolase fold: Y111-I390 Lipases, serine active site: A146-F195 LIPASE HYDROLASE PRECURSOR SIGNAL LIPID DEGRADATION PROTEIN GLYCOPROTEIN ESTERASE TRIACYLGLYCEROL PD003556:N28-L397 TRIACYLGLYCEROL LIPASE, LINGUAL DM02342  P04634 32-394:M33-E395  JC4017 1-394:M1-E395  P07098 35-395:I35-E395  P38571 3-397:N28-Y396 Lipases, serine active site: L166-T175	HMMER SPSCAN HMMER_PFAM PROFILES SCAN BLAST_PRODOR
5	7481764CD1	410	S111 S248 S265 S32 S396 S46 S81 T28	N171 N272	Lipolytic enzymes "G-D-X BL01173A:I118-S130 BL01173B:L149-H175	MOTIFS BLIMPS_BLOCKS
6	72023748CD1	1619	S1008 S1012 S1068 S1132 S1154 S1162 S1169 S1181 S1188 S1195 S1253 S1285 S1485 S1522 S1536 S1584 S184 S278 S343 S430 S531 S543 S576 S606 S65 S678 S760 S801 S875 S89 S970 T1029 T1037 T1081 T1219 T1416	N1059 N815 N828 N997	Phosphatidylinositol-specific phospholipase PI-PLC-X:D848-K993 PI-PLC-Y:A1225-C1340  C2 domain: L1360-T1452	HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6			T1525 T171 T242 T263 T305 T358 T586 T665 T761 T8 T912 Y1232		pleckstrin homology domain (PH):A496-A603 Phosphatidylinositol-specific phospholipase X-box domain proteins BL50007:F1273-G1314, D1439-I1475, L853-G898, T912-Q949, L977-K993 Phospholipase C signature PR00390:P852-Q870, D878-G898, L1453-R1463, T976-K993, L1278-W1299, W1299-L1317 C2 domain signature PR00360:R1381-I1393, N1411-M1424, V1433-D1441 PHOSPHOLIPASE C PHOSPHODIESTERASE 1 PHOSPHATIDYL INOSITOL-4,5-BISPHOSPHATE LIPID DEGRADATION PHOSPHOINOSITIDE SPECIFIC OR CALCIUM BINDING PD001214:D848-K993 PD001202:L1226-P1336 PD004439:Q736-Q847, C488-R703, L821-Q851 1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE D DM00855  P51178 64-472:D742-D1019, S540-R616, P635-A708, L275-W285  P08487 71-500:L735-G1025, I539-M602, G643-R697  P16885 63-486:H743-E1010, I539-L601, W661-H707  A53970 67-522:Y772-P995, I539-M602	HMMER_PFAM BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PRINTS BLAST_PRODOR BLAST_DOMO

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
7	7479063CB1	1228	1-331, 390- 1228	71075936V1 FL7479063_g7230085_000 016_g2204113	896 1	1228 1212
8	7479064CB1	1143	1-264, 928- 1143, 317- 357	7989362H1 (UTRCDIC01) FL7479064_g7230085_000 012_g14749	1 274	696 1143
9	7478088CB1	784	1-784	55146716T1 55146715H1	111 1	784 690
10	7480935CB1	1473	1-264, 928- 1077, 317- 357, 1140- 1473	FL7480935_g9801030_000 005_g758064 7989362H1 (UTRCDIC01)	274 1	1473 696
11	7481764CB1	1619	1-1158	GNN.g9795122_000003.ed it	1	1268
12	72023748CB1	6551	4004-4191, 1861-1978, 3239-3517, 1-1794, 5526-6125, 4754-5258, 2111-2207	g5369512 6875328H1 (EPIMUNN04) 72334958V1 7314090H1 (LIVRNOE07) 55057395J1 6862806F8 (BRAGNON02) 71870033V1 55076578H1 72335316V1 6145149H1 (BRANDIT03) 7617171H1 (KIDNTUE01) 6892787J1 (BRAITDR03) 7019655H1 (PANCNON03) 72334872V1 6935553H1 (SINTPMR02) GNN.g8039388.edit	1204 5670 3898 1509 2908 2462 3455 2166 4535 5453 4153 6154 5899 2652 5237 1	1619 6146 4641 1861 3239 3049 3947 2545 5323 6027 4695 6551 6547 3238 5597 6551

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
7	7479063CB1	MONOTXS05
8	7479064CB1	UTRCDIC01
9	7478088CB1	BRNDIT01
10	7480935CB1	UTRCDIC01
12	72023748CB1	PITUNON01

Table 6

Library	Vector	Library Description
BRONDIT01	pINCY	Library was constructed using RNA isolated from right lower lobe bronchial tissue removed from a pool of 3 asthmatic Caucasian male and female donors, 22- to 51-years-old during bronchial pinch biopsies. Patient history included atopy as determined by positive skin tests to common aero-allergens.
MONOTXS05	pINCY	Subtracted, treated monocyte tissue library was constructed using 7.5 million clones from a treated monocyte library and were subjected to two rounds of subtraction hybridization with 1.03 x 10 <sup>6</sup> clones from a second treated monocyte library. The starting library for subtraction was constructed using treated monocytes from peripheral blood obtained from a 42-year-old female. The cells were treated with anti-interleukin-10 (anti-IL-10) and lipopolysaccharide (LPS). The anti-IL-10 was added at time 0 at 10 ng/ml and LPS was added at 1 hour at 5 ng/ml. The monocytes were isolated from buffy coat by adherence to plastic. Incubation time was 24 hours. The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from monocyte tissue, treated with interleukin-10 (IL10) and lipopolysaccharide (LPS) from the same donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. NAR (1991) 19:1954 and Bonaldo, et al. Genome Research (1996) 6:791.
PITUNON01	pINCY	This normalized pituitary gland tissue library was constructed from 6.92 million independent clones from a pituitary gland tissue library. Starting RNA was made from pituitary gland tissue removed from a 55-year-old male who died from chronic obstructive pulmonary disease. Neuropathology indicated there were no gross abnormalities, other than mild ventricular enlargement. There was no apparent microscopic abnormality in any of the neocortical areas examined, except for a number of silver positive neurons with apical dendrite staining, particularly in the frontal lobe. The significance of this was undetermined. The only other microscopic abnormality was that there was prominent silver staining with some swollen axons in the CA3 region of the anterior and posterior hippocampus. Microscopic sections of the cerebellum revealed mild Bergmann's gliosis in the Purkinje cell layer. Patient history included schizophrenia. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.



Table 6 (cont.)

Library	Vector	Library Description
UTRCDIC01	PSPORT1	<p>This large size fractionated library was constructed using RNA isolated from uterine cervix tissue removed from a 29-year-old Caucasian female during a vaginal hysterectomy and cystocele repair. Pathology indicated the cervix showed mild chronic cervicitis with focal squamous metaplasia. Pathology for the matched tumor tissue indicated intramural uterine leiomyoma. Patient history included hypothyroidism, pelvic floor relaxation, paraplegia, and self catheterization. Previous surgeries included a normal delivery, a laminectomy, and a rhinoplasty. Patient medications included Synthroid. Family history included benign hypertension in the father; and type II diabetes and hyperlipidemia in the mother.</p>

Table 7

Program	Description	Reference	Parameter Threshold
ABI/FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) <i>J. Mol. Biol.</i> 215:403-410; Altschul, S.F. et al. (1997) <i>Nucleic Acids Res.</i> 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less <i>Full Length sequences</i> : Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) <i>Proc. Natl. Acad. Sci. USA</i> 85:2444-2448; Pearson, W.R. (1990) <i>Methods Enzymol.</i> 183:63-98; and Smith, T.F. and M.S. Waterman (1981) <i>Adv. Appl. Math.</i> 2:482-489.	ESTs: fasta E value=1.06E-6 <i>Assembled ESTs</i> : fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less <i>Full Length sequences</i> : fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) <i>Nucleic Acids Res.</i> 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) <i>Methods Enzymol.</i> 266:88-105; and Attwood, T.K. et al. (1997) <i>J. Chem. Inf. Comput. Sci.</i> 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) <i>J. Mol. Biol.</i> 235:1501-1531; Somnhammer, E.L.L. et al. (1988) <i>Nucleic Acids Res.</i> 26:320-322; Durbin, R. et al. (1998) <i>Our World View</i> , in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less <i>Signal peptide hits</i> : Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribkov, M. et al. (1988) CABIOS 4:61-66; Gribkov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairach, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairach, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of
  - 5 SEQ ID NO:1-6,
  - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6,
  - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, and
  - 10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-6.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:7-
- 20 12.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of producing a polypeptide of claim 1, the method comprising:
  - 30 a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
  - b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide selected from the group consisting of:
- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting
  - 5 of SEQ ID NO:7-12,
  - b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:7-12,
  - c) a polynucleotide complementary to a polynucleotide of a),
  - d) a polynucleotide complementary to a polynucleotide of b), and
  - 10 e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.
13. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex
  - 20 is formed between said probe and said target polynucleotide or fragments thereof, and
  - b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.
15. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
  - 30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.
16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable

excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

5

18. A method for treating a disease or condition associated with decreased expression of functional LMM, comprising administering to a patient in need of such treatment the composition of claim 16.

10 19. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

15 20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional LMM, comprising administering to a patient in need of such treatment a composition of  
20 claim 20.

22. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- 25 b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

30 24. A method for treating a disease or condition associated with overexpression of functional LMM, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim

1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a  
5 compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions  
10 permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound,  
and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change  
15 in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method  
20 comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts  
25 of the compound and in the absence of the compound.

28. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at  
30 least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

5

29. A diagnostic test for a condition or disease associated with the expression of LMM in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

10 b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

a) a chimeric antibody,

15 b) a single chain antibody,

c) a Fab fragment,

d) a F(ab')<sub>2</sub> fragment, or

e) a humanized antibody.

20 31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of LMM in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

25 33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of LMM in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

30 35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10, the method comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, or an immunogenic fragment thereof, under conditions to elicit an



antibody response,

- b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10, the method comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

43. A method of detecting a polypeptide having an amino acid sequence selected from the

group consisting of SEQ ID NO:1-6 in a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide  
5 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding  
10 of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-6.

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

15

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

20

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

25

51. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:7.

52. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:8.

30

53. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:9.

54. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:10.

55. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:11.

56. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:12.

<110> INCYTE GENOMICS, INC.

GRIFFIN, Jennifer A.

PATTERSON, Chandra

GANDHI, Ameena R.

LU, Yan

YAO, Monique G.

BAUGHN, Mariah R.

WALIA, Narinder K.

HAFALIA, April J.A.

DING, Li

TRIBOULEY, Catherine M.

DAS, Debopriya

THORNTON, Michael

LAL, Preeti

<120> LIPID METABOLIZING ENZYMES

<130> PI-0203 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/227,429; 60/231,370; 60/233,212; 60/236,885

<151> 2000-08-23; 2000-09-08; 2000-09-15; 2000-09-29

<160> 12

<170> PERL Program

<210> 1

<211> 403

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7479063CD1

<400> i

Met	Pro	Gly	Ile	Ser	Met	Met	Trp	Leu	Leu	Leu	Thr	Thr	Thr	Cys
1				5				10						15
Leu	Ile	Cys	Gly	Thr	Leu	Asn	Ala	Gly	Gly	Phe	Leu	Asp	Leu	Glu
			20					25						30
Asn	Glu	Val	Asn	Pro	Glu	Val	Trp	Met	Asn	Thr	Ser	Glu	Ile	Ile
			35					40						45
Ile	Tyr	Asn	Gly	Tyr	Pro	Ser	Glu	Glu	Tyr	Glu	Val	Thr	Thr	Glu
			50					55						60
Asp	Gly	Tyr	Ile	Leu	Leu	Val	Asn	Arg	Ile	Pro	Tyr	Gly	Arg	Thr
			65					70						75
His	Ala	Arg	Ser	Thr	Gly	Pro	Arg	Pro	Val	Val	Tyr	Met	Gln	His
			80					85						90
Ala	Leu	Phe	Ala	Gly	Gln	Ala	Tyr	Trp	Leu	Glu	Asn	Tyr	Ala	Asn
			95					100						105
Gly	Ser	Leu	Gly	Phe	Leu	Leu	Ala	Asp	Ala	Gly	Tyr	Asp	Val	Trp
			110					115						120
Met	Gly	Asn	Ser	Arg	Gly	Asn	Thr	Trp	Ser	Arg	Arg	His	Lys	Thr

125	130	135
Leu Ser Glu Thr Asp Glu Lys Phe Trp	Ala Phe Ser Phe Asp Glu	
140	145	150
Met Ala Lys Tyr Asp Leu Pro Gly Val	Ile Asp Phe Ile Val Asn	
155	160	165
Lys Thr Gly Gln Glu Lys Leu Tyr Phe	Ile Gly His Ser Leu Gly	
170	175	180
Thr Thr Ile Gly Phe Val Ala Phe Ser	Thr Met Pro Glu Leu Ala	
185	190	195
Gln Arg Ile Lys Met Asn Phe Ala Leu	Gly Pro Thr Ile Ser Phe	
200	205	210
Lys Tyr Pro Thr Gly Ile Phe Thr Arg	Phe Phe Leu Leu Pro Asn	
215	220	225
Ser Ile Ile Lys Ala Val Phe Gly Thr	Lys Gly Phe Phe Leu Glu	
230	235	240
Asp Lys Lys Thr Lys Ile Ala Ser Thr	Lys Ile Cys Asn Asn Lys	
245	250	255
Ile Leu Trp Leu Ile Cys Ser Glu Phe	Met Ser Leu Trp Ala Gly	
260	265	270
Ser Asn Lys Lys Asn Met Asn Gln Ser	Arg Met Asp Val Tyr Met	
275	280	285
Ser His Ala Pro Thr Gly Ser Ser Val	His Asn Ile Leu His Met	
290	295	300
Asn Gln Leu Tyr His Ser Asp Glu Phe	Arg Ala Tyr Asp Trp Gly	
305	310	315
Asn Asp Ala Asp Asn Met Lys His Tyr	Asn Gln Ser His Pro Pro	
320	325	330
Ile Tyr Asp Leu Thr Ala Met Lys Val	Pro Thr Ala Ile Trp Ala	
335	340	345
Gly Gly His Asp Val Leu Val Thr Pro	Gln Asp Val Ala Arg Ile	
350	355	360
Leu Pro Gln Ile Lys Ser Leu His Tyr	Phe Lys Leu Leu Pro Asp	
365	370	375
Trp Asn His Phe Asp Phe Val Trp Gly	Leu Asp Ala Pro Gln Arg	
380	385	390
Met Tyr Ser Glu Ile Ile Ala Leu Met	Lys Ala Tyr Ser	
395	400	

&lt;210&gt; 2

&lt;211&gt; 289

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7479064CD1

&lt;400&gt; 2

Met Trp Gln Leu Leu Ala Ala Ala Cys Trp Met Leu Leu Leu Gly	
1	5 10 15
Ser Met Tyr Gly Tyr Asp Lys Lys Gly Asn Asn Ala Asn Pro Glu	
20	25 30
Ala Asn Met Asn Ile Ser Gln Ile Ile Ser Tyr Trp Gly Tyr Pro	
35	40 45
Tyr Glu Glu Tyr Asp Val Thr Thr Lys Asp Gly Tyr Ile Leu Gly	
50	55 60

```

Ile Tyr Arg Ile Pro His Gly Arg Gly Cys Pro Gly Arg Thr Ala
      65                      70                      75
Pro Lys Pro Ala Val Tyr Leu Gln His Gly Leu Ile Ala Ser Ala
      80                      85                      90
Ser Asn Trp Ile Cys Asn Leu Pro Asn Asn Ser Leu Ala Phe Leu
      95                      100                     105
Leu Ala Asp Ser Gly Tyr Asp Val Trp Leu Gly Asn Ser Arg Gly
      110                     115                     120
Asn Thr Trp Ser Arg Lys His Leu Lys Leu Ser Pro Lys Ser Pro
      125                     130                     135
Glu Tyr Trp Ala Phe Ser Leu Asp Glu Met Ala Lys Tyr Asp Leu
      140                     145                     150
Pro Ala Thr Ile Asn Phe Ile Ile Glu Lys Thr Gly Gln Lys Arg
      155                     160                     165
Leu Tyr Tyr Val Gly His Ser Gln Gly Thr Thr Ile Ala Phe Ile
      170                     175                     180
Ala Phe Ser Thr Asn Pro Glu Leu Ala Lys Lys Ile Lys Ile Phe
      185                     190                     195
Phe Ala Leu Ala Pro Val Val Thr Val Lys Tyr Thr Gln Ser Pro
      200                     205                     210
Met Lys Lys Leu Thr Thr Leu Ser Arg Arg Val Val Lys Val Leu
      215                     220                     225
Phe Gly Asp Lys Met Phe His Pro His Thr Leu Phe Asp Gln Phe
      230                     235                     240
Ile Ala Thr Lys Val Cys Asn Arg Lys Leu Phe Arg Arg Ile Cys
      245                     250                     255
Ser Asn Phe Leu Phe Thr Leu Ser Gly Phe Asp Pro Gln Asn Leu
      260                     265                     270
Asn Met Val Gly Val Ser Asn Trp Val Trp Glu Asn Ile Cys Phe
      275                     280                     285
Val Ala Gln Arg

```

&lt;210&gt; 3

&lt;211&gt; 168

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7478088CD1

&lt;400&gt; 3

```

Met Gln His Met Lys Ser Lys Lys Ile Leu Asp Phe Ala Ala Asn
  1                      5                      10                      15
Val Trp Leu Ser Arg Ile Gln Ala Asp Gly Asp Val Val Cys Glu
      20                      25                      30
Leu Pro Val Val Lys Gly Gly Gln Ala Ile Phe Pro Leu Val Arg
      35                      40                      45
Tyr Gln Val Asp Val Tyr Thr Gly Gln Leu Lys His Ala Glu Thr
      50                      55                      60
Glu Ser Glu Val Phe Leu Cys Leu Phe Gly Glu Arg Gly Asp Ser
      65                      70                      75
Gly Leu Arg Gln Leu Tyr Lys Ser Asn Met Gln Val Lys Phe Gln
      80                      85                      90
Arg Gly Gln Ile Asp Lys Phe Gln Val Ala Ala Val Ser Leu Gly

```

	95		100		105
Lys Leu Gln Lys Val Leu Leu Arg Cys Glu Ala Ser Asp Lys Ser					
	110		115		120
Gln Tyr Trp Tyr Cys Glu Gln Val Ile Val Arg Glu Pro Gly Thr					
	125		130		135
Thr Ser Glu Ser Ile Phe Thr Cys Gln Arg Trp Leu Pro Phe Met					
	140		145		150
Ser Gln Gly Ile Ile His Ser Glu Ile Glu Leu Tyr Leu Gln Gly					
	155		160		165
Glu Asp Tyr					

&lt;210&gt; 4

&lt;211&gt; 399

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7480935CD1

&lt;400&gt; 4

Met Trp Gln Leu Leu Ala Ala Ala Cys Trp Met Leu Leu Leu Gly		
1	5	10
Ser Met Tyr Gly Tyr Asp Lys Lys Gly Asn Asn Ala Asn Pro Glu		
	20	25
Ala Asn Met Asn Ile Ser Gln Ile Ile Ser Tyr Trp Gly Tyr Pro		
	35	40
Tyr Glu Glu Tyr Asp Val Thr Thr Lys Asp Gly Tyr Ile Leu Gly		
	50	55
Ile Tyr Arg Ile Pro His Gly Arg Gly Cys Pro Gly Arg Thr Ala		
	65	70
Pro Lys Pro Ala Val Tyr Leu Gln His Gly Leu Ile Ala Ser Ala		
	80	85
Ser Asn Trp Ile Cys Asn Leu Pro Asn Asn Ser Leu Ala Phe Leu		
	95	100
Leu Ala Asp Ser Gly Tyr Asp Val Trp Leu Gly Asn Ser Arg Gly		
	110	115
Asn Thr Trp Ser Arg Lys His Leu Lys Leu Ser Pro Lys Ser Pro		
	125	130
Glu Tyr Trp Ala Phe Ser Leu Asp Glu Met Ala Lys Tyr Asp Leu		
	140	145
Pro Ala Thr Ile Asn Phe Ile Ile Glu Lys Thr Gly Gln Lys Arg		
	155	160
Leu Tyr Tyr Val Gly His Ser Gln Gly Thr Thr Ile Ala Phe Ile		
	170	175
Ala Phe Ser Thr Asn Pro Glu Leu Ala Lys Lys Ile Lys Ile Phe		
	185	190
Phe Ala Leu Ala Pro Val Val Thr Val Lys Tyr Thr Gln Ser Pro		
	200	205
Met Lys Lys Leu Thr Thr Leu Ser Arg Arg Val Val Lys Val Leu		
	215	220
Phe Gly Asp Lys Met Phe His Pro His Thr Leu Phe Asp Gln Phe		
	230	235
Ile Ala Thr Lys Val Cys Asn Arg Lys Leu Phe Arg Arg Ile Cys		
	245	250
		255

Ser Asn Phe Leu Phe Thr Leu Ser Gly Phe Asp Pro Gln Asn Leu  
 260 265 270  
 Asn Met Ser Arg Leu Asp Val Tyr Leu Ser His Asn Pro Ala Gly  
 275 280 285  
 Thr Ser Val Gln Asn Met Leu His Trp Ala Gln Ala Val Asn Ser  
 290 295 300  
 Gly Gln Leu Gln Ala Phe Asp Trp Gly Asn Ser Asp Gln Asn Met  
 305 310 315  
 Met His Phe His Gln Leu Thr Pro Pro Leu Tyr Asn Ile Thr Lys  
 320 325 330  
 Met Glu Val Pro Thr Ala Ile Trp Asn Gly Gly Gln Asp Ile Val  
 335 340 345  
 Ala Asp Pro Lys Asp Val Glu Asn Leu Leu Pro Gln Ile Ala Asn  
 350 355 360  
 Leu Ile Tyr Tyr Lys Leu Ile Pro His Tyr Asn His Val Asp Phe  
 365 370 375  
 Tyr Leu Gly Glu Asp Ala Pro Gln Glu Ile Tyr Gln Asp Leu Ile  
 380 385 390  
 Ile Leu Met Glu Glu Tyr Leu Gln Asn  
 395

&lt;210&gt; 5

&lt;211&gt; 410

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7481764CD1

&lt;400&gt; 5

Met Ala Lys Ser Val Glu Gln Leu Pro Trp Ala Arg His Cys Ala  
 1 5 10 15  
 Glu Tyr Met Asp Met Phe Ser Ser Leu Ile Leu Thr Thr Pro Arg  
 20 25 30  
 Val Ser Pro Lys Leu Gly Tyr Pro Leu Ala Asn Ser His His Tyr  
 35 40 45  
 Ser Ile Lys Ser Leu Trp Gly Glu Lys Leu Glu Asn Pro Ala Leu  
 50 55 60  
 Tyr Leu Asp Thr Val Gln Ser Leu Trp Ile Pro Glu Glu Pro Pro  
 65 70 75  
 Val Pro Thr Gly Gly Ser Val Arg Ile Lys Lys Asp Pro Glu Leu  
 80 85 90  
 Val Val Thr Asp Ser Arg Phe Gly Thr Ile Pro Val Arg Leu Phe  
 95 100 105  
 Gln Pro Lys Ala Ala Ser Ser Arg Pro Arg Arg Gly Ile Ile Phe  
 110 115 120  
 Tyr His Gly Gly Ala Thr Val Phe Gly Ser Leu Asp Cys Tyr His  
 125 130 135  
 Gly Leu Cys Asn Tyr Leu Ala Arg Glu Thr Glu Ser Val Leu Leu  
 140 145 150  
 Met Ile Gly Tyr Arg Lys Leu Pro Asp His His Ser Pro Ala Leu  
 155 160 165  
 Phe Gln Asp Cys Met Asn Ala Ser Ile His Phe Leu Lys Ala Leu  
 170 175 180  
 Glu Thr Tyr Gly Val Asp Pro Ser Arg Val Val Val Cys Gly Glu



	185		190		195
Ser Val Gly Gly	Ala Ala Val Ala Ala	Ile Thr Gln Ala Leu Val			
	200		205		210
Gly Arg Ser Asp	Leu Pro Arg Ile Arg	Ala Gln Val Leu Ile Tyr			
	215		220		225
Pro Val Val Gln	Ala Phe Cys Leu Gln	Leu Pro Ser Phe Gln Gln			
	230		235		240
Asn Gln Asn Val	Pro Leu Leu Ser Arg	Lys Phe Met Val Thr Ser			
	245		250		255
Leu Cys Asn Tyr	Leu Ala Ile Asp Leu	Ser Trp Arg Asp Ala Ile			
	260		265		270
Leu Asn Gly Thr	Cys Val Pro Pro Asp	Val Trp Arg Lys Tyr Glu			
	275		280		285
Lys Trp Leu Ser	Pro Asp Asn Ile Pro	Lys Lys Phe Lys Asn Arg			
	290		295		300
Gly Tyr Gln Pro	Trp Ser Pro Gly Pro	Phe Asn Glu Ala Ala Tyr			
	305		310		315
Leu Glu Ala Lys	His Met Leu Asp Val	Glu Asn Ser Pro Leu Ile			
	320		325		330
Ala Asp Asp Glu	Val Ile Ala Gln Leu	Pro Glu Ala Phe Leu Val			
	335		340		345
Ser Cys Glu Asn	Asp Ile Leu Arg Asp	Asp Ser Leu Leu Tyr Lys			
	350		355		360
Lys Arg Leu Glu	Asp Gln Gly Val Arg	Val Thr Trp Tyr His Leu			
	365		370		375
Tyr Asp Gly Phe	His Gly Ser Ile Ile	Phe Phe Asp Lys Lys Ala			
	380		385		390
Leu Ser Phe Pro	Cys Ser Leu Lys Ile	Val Asn Ala Val Val Ser			
	395		400		405
Tyr Ile Lys Gly	Ile				
	410				

&lt;210&gt; 6

&lt;211&gt; 1619

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 72023748CD1

&lt;400&gt; 6

Met Ala Cys Cys	Gln Val Val Thr Pro	Arg His Leu Gly Pro Ser
1	5	10 15
Gly Ala Thr Thr	Val Ala Leu Val Leu	Val Gly Arg Leu Gln Ser
	20	25 30
Pro Cys Leu Leu	Lys Phe Ala Glu Arg	Gly Cys Ser Arg Glu Ser
	35	40 45
Val Tyr Cys Ser	Arg Ser Ser Gly Ser	Val Ser Leu Glu Thr Tyr
	50	55 60
Val Gly Gln Asn	Ser Glu Ala Glu Thr	Glu Ala Pro His Glu Gly
	65	70 75
Pro Thr Pro Leu	Ala Gly Gly Gly Phe	Ser Gly Leu Cys Ser Trp
	80	85 90
Arg Leu Leu Gly	Asp Ser Cys Ser Pro	Glu Thr Val Ala Phe Ser
	95	100 105

Gly Thr His Ser Gly His Pro Leu Leu Ala Ala Pro Glu Asp Met	110	115	120
Val Leu Gly Phe Cys Ala Gly Ser Val Trp Gly Cys Asp Arg His	125	130	135
Gly Leu Cys Ala Arg Gly Ala Pro Ala Arg Ala Gly Ala Gln Gly	140	145	150
Thr Pro Gly Ala Ser Thr Gly Pro Gly Val Pro Arg Ala Val Pro	155	160	165
Val Thr Gly Ile Gly Thr Arg Arg Pro Trp Lys Ser Leu Gly Pro	170	175	180
Gln Val Gly Ser Ala Lys Thr Lys Pro Leu Ile Glu Val Gly Lys	185	190	195
Gly Gly Pro Glu Leu Gln Ser Ala Pro Trp Thr Gly Ala Ala Ala	200	205	210
Ala Phe Ser Gly Gly Pro Ala Val Ala Val Arg Pro Gln Gly Ser	215	220	225
Leu Glu Leu His Ser Glu Gly Gly Glu Pro Ala Glu Gln Ser Leu	230	235	240
Gln Thr Pro Arg Ala Ala Glu Asp Gly Ala Asp Gly Pro Val Ser	245	250	255
Trp Val Gln Pro Ser Ile Cys Thr Ala Asn Glu Leu Met Thr Met	260	265	270
Ala Ser Gln Phe Leu Ala Pro Ser Pro Gly Glu Ala Met Gly Trp	275	280	285
Pro Cys Pro Gly Ser Ala Leu Cys Trp Gly Ser Trp Gln Ala Arg	290	295	300
Leu His Leu Val Thr Phe Gln Glu Arg Leu Pro His Arg Gly Cys	305	310	315
His Glu Pro Gly Ser Leu Cys Ser Ser Ser Gly Val Ser Trp Thr	320	325	330
Val Ala Gly Gly Lys Ala Val His Leu Pro Ser Glu Ser Pro Arg	335	340	345
Pro Ala Ser Ala Cys Tyr Val Leu Asp His Gly Leu Thr Gly Thr	350	355	360
Glu Gln Gly Gly Gly Leu Arg Gly Pro Cys Pro Arg Asp Val Pro	365	370	375
Gly Ser Gly Ser Ala Leu Arg Thr Gly Glu Pro His Cys Ala Val	380	385	390
Gln Gly Pro Gly Gly Ala Ala Gly Gln Asn Val Gly Thr Asn Val	395	400	405
Pro Trp Ala Trp Gly His Leu Leu Arg His Ser Ser Val Thr Val	410	415	420
Glu Lys Gly Arg Trp Ala Ser Val Ala Ser Val Lys Gln Ala Arg	425	430	435
Leu Ser Ser Gly His Val Trp Ser Met Ala Leu Pro Arg Gln Pro	440	445	450
Asp Gln Gly Asn Gly Gly Leu Ala Gly Gly Gly Val Ala Val Pro	455	460	465
Pro Leu Thr Ala Val Ser Leu Leu Arg Gly Val Ala Val Pro Pro	470	475	480
Leu Thr Ala Val Ser Leu Pro Cys Pro Val Glu Arg Cys Met Gly	485	490	495
Ala Met Gln Glu Gly Met Gln Met Val Lys Leu Arg Gly Gly Ser	500	505	510
Lys Gly Leu Val Arg Phe Tyr Tyr Leu Asp Glu His Arg Ser Cys	515	520	525

Ile Arg Trp Arg	Pro Ser Arg Lys Asn	Glu Lys Ala Lys Ile Ser	
	530	535	540
Ile Asp Ser Ile	Gln Glu Val Ser Glu	Gly Arg Gln Ser Glu Val	
	545	550	555
Phe Gln Arg Tyr	Pro Asp Gly Ser Phe	Asp Pro Asn Cys Cys Phe	
	560	565	570
Ser Ile Tyr His	Gly Ser His Arg Glu	Ser Leu Asp Leu Val Ser	
	575	580	585
Thr Ser Ser Glu	Val Ala Arg Thr Trp	Val Thr Gly Leu Arg Tyr	
	590	595	600
Leu Met Ala Gly	Ile Ser Asp Glu Asp	Ser Leu Ala Arg Arg Gln	
	605	610	615
Arg Thr Arg Asp	Gln Tyr Pro Trp Ala	Pro Ile Gly Gln Cys Arg	
	620	625	630
Pro Arg Asp Arg	Pro Leu Gly Cys Ser	Pro Trp Gly Gly Leu Ser	
	635	640	645
Phe Ala Gly Ser	His Thr Gly Glu Val	Ala Gly Gln Arg Val Glu	
	650	655	660
Trp Leu Lys Gln	Thr Phe Asp Glu Ala	Asp Lys Asn Gly Asp Gly	
	665	670	675
Ser Leu Ser Ile	Gly Glu Val Leu Gln	Leu Leu His Lys Leu Asn	
	680	685	690
Val Asn Leu Pro	Arg Gln Arg Val Lys	Gln Met Phe Arg Val Ala	
	695	700	705
Gly His Ala Trp	Leu Glu Gln Gly Lys	Leu Ala Cys Ser Gln Asp	
	710	715	720
Arg Ala Leu Val	Glu Val Pro Met Gly	Thr Gln Gly Leu Ala Leu	
	725	730	735
Gln Glu Ala Asp	Thr Asp Asp His Gln	Gly Thr Leu Gly Phe Glu	
	740	745	750
Glu Phe Cys Ala	Phe Tyr Lys Met Met	Ser Thr Arg Arg Asp Leu	
	755	760	765
Tyr Leu Leu Met	Leu Thr Tyr Ser Asn	His Lys Asp His Leu Asp	
	770	775	780
Ala Ala Ser Leu	Gln Arg Phe Leu Gln	Val Glu Gln Lys Met Ala	
	785	790	795
Gly Val Thr Leu	Glu Ser Cys Gln Asp	Ile Ile Glu Gln Phe Glu	
	800	805	810
Pro Cys Pro Glu	Asn Lys Ser Lys Gly	Leu Leu Gly Ile Asp Gly	
	815	820	825
Phe Thr Asn Tyr	Thr Arg Ser Pro Ala	Gly Asp Ile Phe Asn Pro	
	830	835	840
Glu His His His	Val His Gln Asp Met	Thr Gln Pro Leu Ser His	
	845	850	855
Tyr Phe Ile Thr	Ser Ser His Asn Thr	Tyr Leu Val Gly Asp Gln	
	860	865	870
Leu Met Ser Gln	Ser Arg Val Asp Met	Tyr Ala Trp Val Leu Gln	
	875	880	885
Ala Gly Cys Arg	Cys Val Glu Val Asp	Cys Trp Asp Gly Pro Asp	
	890	895	900
Gly Glu Pro Ile	Val His His Gly Tyr	Thr Leu Thr Ser Lys Ile	
	905	910	915
Leu Phe Lys Asp	Val Ile Glu Thr Ile	Asn Lys Tyr Ala Phe Ile	
	920	925	930
Lys Asn Glu Tyr	Pro Val Ile Leu Ser	Ile Glu Asn His Cys Ser	
	935	940	945

9/15

Ser Gly Gln Gln Leu Pro Lys Pro Arg Asp Ser Met Leu Gly Asp  
 1370 1375 1380  
 Arg Gly Glu Ile Ile Asp Pro Phe Val Glu Val Glu Ile Ile Gly  
 1385 1390 1395  
 Leu Pro Val Asp Cys Ser Arg Glu Gln Thr Arg Val Val Asp Asp  
 1400 1405 1410  
 Asn Gly Phe Asn Pro Thr Trp Glu Glu Thr Leu Val Phe Met Val  
 1415 1420 1425  
 His Met Pro Glu Ile Ala Leu Val Arg Phe Leu Val Trp Asp His  
 1430 1435 1440  
 Asp Pro Ile Gly Arg Asp Phe Ile Gly Gln Arg Thr Leu Ala Phe  
 1445 1450 1455  
 Ser Ser Met Met Pro Gly Tyr Arg His Val Tyr Leu Glu Gly Met  
 1460 1465 1470  
 Glu Glu Ala Ser Ile Phe Val His Val Ala Val Ser Asp Ile Ser  
 1475 1480 1485  
 Gly Lys Val Lys Gln Ala Leu Gly Leu Lys Gly Leu Phe Leu Arg  
 1490 1495 1500  
 Gly Pro Lys Pro Gly Ser Leu Asp Ser His Ala Ala Gly Arg Pro  
 1505 1510 1515  
 Pro Ala Arg Pro Ser Val Ser Gln Arg Thr Leu Arg Arg Thr Ala  
 1520 1525 1530  
 Ser Ala Pro Thr Lys Ser Gln Lys Pro Gly Arg Arg Gly Phe Pro  
 1535 1540 1545  
 Glu Leu Val Leu Gly Thr Arg Asp Thr Gly Ser Lys Gly Val Ala  
 1550 1555 1560  
 Asp Asp Val Val Pro Pro Gly Pro Gly Pro Ala Pro Glu Ala Pro  
 1565 1570 1575  
 Ala Gln Glu Gly Pro Gly Ser Gly Ser Pro Arg Gly Lys Ala Pro  
 1580 1585 1590  
 Ala Ala Val Ala Glu Lys Ser Pro Val Arg Val Arg Pro Cys Val  
 1595 1600 1605  
 Ser Trp Thr Ala Pro Gly Leu Leu Gly Trp Pro Pro His Ala  
 1610 1615

&lt;210&gt; 7

&lt;211&gt; 1228

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7479063CB1

&lt;400&gt; 7

atgccaggca tttctatgat gtggctgctt ttaacaacaa cttgtttgat ctgtggaact 60  
 ttaaattgctg gtggattcct tgatttggaa aatgaagtga atcctgaggt gtggatgaat 120  
 actagtgaat tcatcatcta caatggctac ccagtggaag agtatgaagt caccactgaa 180  
 gatgggtata tactccttgt caacagaatt ccttatgggc gaacacatgc taggagcaca 240  
 ggtccccggc cagtttgtga tatgcagcat gccctgtttg caggacaagc ctactggctt 300  
 gagaattatg ccaatggaag ccttggattc cttctagcag atgcaggtta tgatgtatgg 360  
 atgggaaaca gtcggggaaa cacttgggtca agaagacaca aaacactctc agagacagat 420  
 gagaaattct gggccttttag ttttgatgaa atggccaaat atgatctccc aggagtaata 480  
 gacttcattg taaataaaac tggtcaggag aaattgtatt tcattggaca ttacttggc 540  
 actacaatag gggtttgtac cttttccacc atgcctgaac tggcacaag aatcaaaatg 600  
 aattttgect tgggtcctac gatctcattc aaatatccca cgggcatttt taccaggttt 660

```

tttctacttc caaattccat aatcaaggct gtttttggtta ccaaagggtt ctttttagaa 720
gataagaaaa cgaagatagc ttctacccaa atctgcaaca ataagatact ctgggttgata 780
tgtagcgaat ttatgtcctt atgggctgga tccaacaaga aaaatatgaa tcagagtcga 840
atggatgtgt atatgtcaca tgctccact gggtcatcag tacacaacat tctgcatatg 900
aatcagcttt accactctga tgaattcaga gcttatgact ggggaaatga cgctgataat 960
atgaaacatt acaatcagag tcatccccct atatatgacc tgactgccat gaaagtgcct 1020
actgctattt gggctgggtg acatgatgtc ctgtaacac cccaggatgt ggccaggata 1080
ctccctcaaa tcaagagtct tcattacttt aagctattgc cagattggaa ccactttgat 1140
tttgtctggg gctctgatgc ccctcaacgg atgtacagtg aaatcatagc tttaatgaag 1200
gcatattcct aaaaaaaaaa aaaaaaaaaa 1228

```

&lt;210&gt; 8

&lt;211&gt; 1143

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7479064CB1

&lt;400&gt; 8

```

gagttgagaa atttctagtt ggcctatctt ttgagctgtt taaaacagag catgttccac 60
tgacagttgc ctctggaagg atctggaatg cacctgaagc tcccaggact tgaagacaga 120
cctccaaaga cttggatcag gttaagcaag atacagcaag agtataatgt gacaacttca 180
gatggagagt cttctagcaa agaattcgaa accgacttgc agaagttaaa tgcccttgga 240
aagggactgc tattttatta aggcagatcc caaatgtggc agcttttagc agcagcatgc 300
tggatgcttc ttcttgatc tatgtatggt tatgacaaga aaggaaacaa tgccaaccct 360
gaagctaata tgaatattag ccagattatt tcttactggg gttatcctta tgaagagtat 420
gatgttacaa caaaagatgg ttatatcctt ggaatttata ggattccaca tggaagagga 480
tgcccaggga ggacagctcc aaagcctgct gtgtatttgc agcatggctt aattgcatct 540
gccagtaact ggatttgcaa cctgcccaac aacagtttgg ctttccctct gccagatagt 600
ggttatgacg tgtggttggg gaacagccga ggaaacactt ggtccagaaa acaccttaaa 660
ttgtcaccga aatcaccaga atactgggcc ttcagtttgg atgagatggc taaatatgac 720
cttcagcca caatcaattt tatcatagag aaaactggac agaagcgact ctactacgtg 780
ggccactcac aaggcaccac catagctttt atagcatttt ctacaaacc agaactggct 840
aaaaagatta agatatTTTT tgactggct ccagttgtca cagttaaata caccctaaagt 900
cctatgaaaa aactaacaac ctttccagg cgagtagtta aggtgttgtt tgggtgacaaa 960
atgttccacc ctcatacatt gtttgaccaa ttcattgcca ccaaagtgtg caatcgaaag 1020
ctattccgtc gtatttgag caacttccta tttactctga gtggatttga tccgcaaac 1080
ttaaatatgg taggtgtaag taattgggtc tgggaaaaca tttgttttgt tgcacagagg 1140
tga 1143

```

&lt;210&gt; 9

&lt;211&gt; 784

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7478088CB1

&lt;400&gt; 9

```

taaaggagtc actgggccaa taagtcttcg caaggacagc tcagagcagc tcttccttcc 60
aggacacgag gatgagtttc aggttgaaat aagaaacact ggaaacatat ataagataag 120
gataggacat gatggaacca gtgagcaacc cgagtggaaac ttgcagaggg tgaccatgca 180
acatatgaaa agcaagaaga tcctagattt tgcagcaaac gtatggctgt ctcggattca 240

```

ggcagatggg gatgttgtct gtgagcttcc tgtagtaaaa ggaggacaag ctatTTTTcc 300  
attagtga taccaggttg acgtctacac tgggcagctg aagcatgcag agacggagtc 360  
tgaggTTTTc ctctgtctct ttggggagag gggagactct ggcctcagac agctgtacaa 420  
atctaactg caagtaaaat ttcaaagagg acagattgat aaatttcaag tagcagcagt 480  
gtcgcttgga aaactgcaga aggtgctgtt gcgctgtgag gccagtgaca aatcacagta 540  
ctggactgt gaacaagtca tagtcagaga gcccggcacc acatccgagt ccatctttac 600  
ctgtcaaaga tggcttcctt tcatgtctca gggcataatt cattcagaaa ttgaacttta 660  
tcttcaaggt gaggattact gaaacatca gtagaagaaa gtacaagatg attttgttac 720  
atctggacat catggaacac tattccccag ttttcctcag aacatccct taatacttat 780  
gttt 784

<210> 10

<211> 1473

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7480935CB1

<400> 10

gagttgagaa atttctagtt gcgccatctt ttgagctgtt taaaacagag catgttccac 60  
tgacagttgc ctctggaagg atctggaatg cacctgaagc tcccaggact tgaagacaga 120  
cctccaaaga cttggatcag gttaagcaag atacagcaag agtataatgt gacaacttca 180  
gatggagagt cttctagcaa agaattcgaa accgacttgc agaagttaaa tgcccttgga 240  
aagggaactgc tattttatta aggcagatcc caaatgtggc agcttttagc agcagcatgc 300  
tggatgcttc ttcttgatc tatgtatggt tatgacaaga aaggaaacaa tgcaaccct 360  
gaagctaata tgaatattag ccagattatt tcttactggg gttatcctta tgaagagtat 420  
gatgttacia caaaaagatgg ttatatcctt ggaatttata ggattccaca tggaagagga 480  
tgcccaggga ggacagctcc aaagcctgct gtgtatttgc agcatggctt aattgcatct 540  
gccagtaact ggatttgcaa cctgcccac aacagtttgg ctttccttct ggcagatagt 600  
ggttatgacg tgtggttggg gaacagccga ggaaacactt ggtccagaaa acaccttaa 660  
ttgtcaccga aatcacccga atactggcc ttcagtttgg atgagatggc taaatatgac 720  
cttcagcca caatcaattt tatcatagag aaaactggac agaagcgact ctactacgtg 780  
ggccactcac aaggcaccac catagctttt atagcatttt ctacaaacc agaactggct 840  
aaaaagatta agatattttt tgcaactggc ccagttgtca cagttaaata caccacaaagt 900  
cctatgaaaa aactaacaac ctttccagg cgagtagtta aggtgttgtt tgggtgacaa 960  
atgttccacc ctcatatcatt gtttgacca ttcatgcca ccaaagtgtg caatcgaaag 1020  
ctattccgtc gtatttgcag caacttccta tttactctga gtggatttga tccgcaaaac 1080  
ttaaatatga gtcgcttgga tgtttatttg tcacacaatc ctgcggaac atctgttcag 1140  
aatatgctgc actgggctca ggctgttaat tctggtcagc tccaagcttt tgattgggga 1200  
aactctgac agaacaatgat gcacttccat cagcttacac ctctttata caacattact 1260  
aagatggaag ttccaacagc aatatggaat ggtggacagg acattgtggc tgatcccaag 1320  
gatgttgaaa atttacttcc tcaaattgct aaccttattt attacaagct gattccacac 1380  
tacaatcatg tggattttta ccttgagag gatgcacctc aggaaattta ccaagacct 1440  
attatattga tggagaata ttacaaaat taa 1473

<210> 11

<211> 1619

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7481764CB1

&lt;400&gt; 11

```
cctccccaca acaacaacag caccaccact atataatggc taaatctgtt gagcagttgc 60
catggggccag acactgtgct gagtacatgg atatgttttc ttctttaatc ctcaaaccc 120
ctcgagtcag ccccaagcta ggctaccctt tggcaaattc acatcattat tcaatcaaga 180
gcctctgggg agaaaagtgt gaaaaccag ccctctacct ggacacagtc cagagcctat 240
ggattcctga agagccccct gtacctacag gaggcagcgt gagaattaaa aaggaccctg 300
aacttgttgt gaccgacctg cgttttgga cgataccctg gaggtgttc cagccgaagg 360
cagcatctc cagaccccg cgaggcatca tctctacca tggaggggac acagtatttg 420
ggagcctgga ttgttaccat ggcctgtgca attatctggc ccgggagact gaatctgtac 480
ttctgatgat tgggtaccgc aagcttcctg accaccattc ccctgccctt ttccaagact 540
gcatgaatgc ctccattcac ttctgaagg ccctggaaac ctatgggggtg gacccctcca 600
gggttgttgt ctgtggagaa agcgtcggag gtgcagcgtt ggccgccatc acccaggcct 660
tgggtggcag atcagatctt ccccgatcc gggctcaggt tctgatttat ccagttgtcc 720
aggcattctg tttgcagttg ccctcttct agcagaacca aaatgtcca ttactttccc 780
ggaagtcat ggtgacttct ctgtgtaact atctggccat tgacctctcc tggcgtgacg 840
ccatcttgaa cggcacttgt gtacccccag acgtctggag gaagtacgag aagtggctca 900
gccctgacaa catccccag aaatttaaga acagaggcta ccaaccctgg tctccggcc 960
cttttaatga agctgcctat ctagaagcca aacatatgct ggatgtagaa aattcacccc 1020
tgatagcaga tgatgaggtc atcgtcagc ttcttgaggc cttctgtgtg agctgtgaga 1080
atgacatact ccgtgatgac agcttgctct ataagaagcg cttggaggac caggggggtcc 1140
gcgtgacatg gtaccacctg tatgatggtt ttcacggatc cattatcttt ttgataaga 1200
aggctctctc tttcccatgt tcctgaaga ttgtgaatgc tgtagtcatg tatataaagg 1260
gcatatgata gtaaccctgg gcccgagga ggaaggggca agtatggact ctaccagaaa 1320
ccgggtgctt tagtgagttc tattttattg actaaagagg tgctacatca atgcttgggg 1380
cagctgggaa ggggtgagaag taagctaaca gtcttgctta gtattcaaga aaatccaaa 1440
tgtgtctgtt tcttccagc actaacaatg tccattgctg gatctagcga cattctctaa 1500
cattcccatt taggtgaaat aaatatcaaa aggagaaaaa aatgccttta aaaatttctc 1560
aaagcccaa catataagat ctgtgcagaa taaatgcaa caactggtca taccgtcaa 1619
```

&lt;210&gt; 12

&lt;211&gt; 6551

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 72023748CB1

&lt;400&gt; 12

```
atggcctgct gtcaggtggt caccccaagg catcttggtc catccggggc aactacggtc 60
gccctggtgc tggtaggcag gctgcagagc ccctgcctgc tgaagtttgc tgaacggggc 120
tgcagccgag aatctgttta ctgcagcagg tcatcgggaa gcgtttcctt ggagacatat 180
gtaggacaga actcggaaagc tgaaaccgag gcacccacg aaggcccccac accgctggcg 240
ggaggcgggt tctcagggct ctgctcctgg aggtctgctg gagacagctg cagcccgag 300
actgtggcat tttctggaac acactcaggg cacccttgc tggcagcccc agaggacatg 360
gtgctgggct tctgtgcggg cagcgtgtgg ggctgcgac ggcatgggct ctgcgcgcg 420
gggtgccctg ctcggtctgg ggctcagggg acgcccgggg ccagcacagg gcccggtgtc 480
cctcggggcg tgccggtaac tgggattggc acccgccggc catggaagag cctgggcccc 540
caggtgggct cagccaagac caagcccctg atcgaggttg gcaaagggtg ccctgagctg 600
cagtgcagac catggacagg ggcagctgct gctttctctg gggggcctgc agttgctgtg 660
cgtcctcagg gttccctgga gttacactct gaggggaggg agccagcaga acagtcccta 720
cagactccac tgccgccga ggacggagca gatgggccag tgctctgggt ccagccctcc 780
atctgcacag ctaatgagct gatgacgat gcgtccagtt tccctggcac ttctcctgga 840
gaagccatgg gctggccctg cccgggaggt gccctttgct ggggcagctg gcaggccagg 900
ctccatctgg tgacttttca ggaaaggctc ccccaccgtg gctgccacga gcctggctcc 960
ctctgcagct cctctggggt gtctggagc gtggctgggg gcaaggccgt gcacctgct 1020
```



tcagagagcc	ccaggccagc	cagcgccctgc	tacgtgttgg	atcatggcct	cacagggaca	1080
gagcagggag	gaggacttcg	aggtccctgc	cccagggatg	tccccgggtc	aggcagtgct	1140
ctgaggactg	gggagcccca	ctgtgctgtg	cagggaacctg	ggggagcagc	agggcagaat	1200
gtggggacaa	acgtgccctg	ggcatggggc	cacctccctcc	gccacagctc	ggtgaccgtg	1260
gagaaaaggcc	ggtgggectc	tgtggcctcc	gtgaagcagg	cccggctgtc	gtcaggccat	1320
gtctgttcca	tggccctccc	ccgacagccg	gaccaaggga	acggtggcct	ggctggcgga	1380
ggagtggccg	tgcctccgct	gacggccgtg	tctctcctgc	gaggagtggc	tgtgcctccg	1440
ctgacagccg	tgtctctccc	gtgtccagt	gagcggtgca	tgggtgccat	gcaagagggg	1500
atgcagatgg	tgaagctgcg	tggcggctcc	aagggcctgg	tccgcttcta	ctacctggac	1560
gagcaccgct	cctgcatccg	ctggaggccc	tcacgcaaga	acgagaaggc	caagatctcc	1620
atcgactcca	tccaggaggt	gagtgaaggg	cggcagtcgg	aggtcttcca	gcgctaccct	1680
gacggcagct	tgcaccccaa	ctgctgcttc	agcatctacc	acggcagcca	ccgcgagtcg	1740
ctggacctgg	tctccaccag	cagcgaggtg	gcgcgcacct	gggtcactgg	cctgcgctac	1800
ctcatggccg	gcatcagcga	cgaggacagc	ctggctcgcc	gccagcgcac	cagggaccaa	1860
tatccttggg	cacctatcgg	gcaatgcaga	cccagggacc	ggcccccttg	ctgctcacc	1920
tggggggggc	tgtcctttgc	cgggtcacac	actggggaag	tggccggcca	gaggggtggag	1980
tggctgaagc	agacgtttga	cgaggccgac	aagaacgggg	atggcagcct	gagcattggc	2040
gaggtcctgc	agctgctgca	caagctcaac	gtgaacctgc	cccggcagag	ggtgaagcag	2100
atgttcaggg	tggctggtca	tgcctggctg	gagcaaggga	agctggcctg	ctcacaggac	2160
agggccctgg	tgcaggtgcc	aatggggacc	caaggccctg	cattgcagga	agcggacacg	2220
gatgaccacc	aagggaagct	gggttttgaa	gagttctgtg	ccttctacaa	gatgatgtcc	2280
acccgcgggg	acctctacct	gctcatgtcg	acctacagca	accacaagga	ccacctggat	2340
gccgccagcc	tgcagcgctt	cctgcaggtg	gagcagaaga	tggcgggtgt	gacctcgag	2400
agctgccagg	acatcatcga	gcagtttgag	ccatgccag	aaaacaagag	taaggggctg	2460
ctgggcattg	atggcttcac	caactacacc	aggagccctg	ctggtgacat	cttcaaccct	2520
gagcaccacc	atgtgcacca	ggacatgacg	cagccgctga	gccactactt	catcacctcg	2580
tcccadaaca	cctacctcgt	gggtgaccag	ctcatgtccc	agtcacgggt	ggacatgtat	2640
gcttgggtcc	tgcaggtcgg	ctgccgctgc	gtggaggtgg	actgctggga	tgggcccagc	2700
ggggagccca	ttgtgcacca	tggctacact	ctgacttcca	agatcctctt	caaagacgtc	2760
attgaaacca	tcaacaaata	tgccttcac	aagaatgagt	accagtgat	cctgtccatc	2820
gaaaaccact	gcagtgtcat	ccagcagaag	aaaatggccc	agtatctgac	tgacatcctt	2880
ggggacaagc	tggacctgtc	atcagtgaag	agtgaagatg	ccaccacact	cccctctcca	2940
cagatgtcca	agggcaagat	cctcgtgaag	gggaagaagc	tcccagccaa	catcagcgag	3000
gatgcggagg	aaggcgaggt	gtctgatgag	gcagtgctg	atgagattga	cgatgactgc	3060
aagctcctca	atggggatgc	atccaccaat	cgaaagcgtg	tagaaaacac	tgctaagagg	3120
aaactggatt	ccctcatcaa	agagtgcgaag	attcgggact	gtgaggaccc	caacaacttc	3180
tccgtctcca	cactgtcccc	atctggaaag	ctcggacgca	aggtagaggc	caaaaagggtg	3240
acacccctga	tgccgacagg	cccccccgac	agccagcctg	tgggccagcc	agggcccccg	3300
aaccgaggca	gtgtccaggc	aggcccaggt	ccaggctggg	gagagcgggc	agcagtcgga	3360
gctgtttgct	gtggcggtgc	tgagcgccga	gtgtctgtgg	tcagtagagc	tacagtggcc	3420
tctttgcagc	agtgtgggag	catccagggg	tgtgggagga	gcaaggctga	agaggacgtg	3480
gagtcgtggg	aggatgccgg	ggccagcaga	cgcaatggcc	gcctcgtcgt	gggaagcttc	3540
tccaggcgca	agaagaaggg	cagcaagctg	aagaaggcgg	ccagcgaggga	ggaggagat	3600
gagggctcagg	actccccggg	aggccagagc	cgaggggcga	cccggcagaa	gaagaccatg	3660
aagctgtccc	gggcccctct	tgacctgggt	aagtacacca	agtcctgggc	caccacagac	3720
atagagatgg	aggcggcgct	cagctggcag	gtgtcgtcct	tcagcgagac	caaggcccac	3780
cagattctgc	agcagaagcc	ggcgcagtac	ctacgcttca	accagcagca	gctctcccgc	3840
atctacccct	cctcctaccg	tgtggactcc	agcaactaca	acccgcagcc	cttctggaac	3900
gccggctgcc	aaatggttgc	cctgaactac	cagtcagagg	ggcggatgct	gcagctgaac	3960
cgagccaagt	tcagcgccaa	cgttggtctg	ggctacgtac	tcaagcctgg	gtgcatgtgc	4020
cagggcggtg	tcaaccccaa	ctcggaggac	cccctgcccg	ggcagctcaa	gaagcagctg	4080
gtgctccgga	tcatcagtg	ccagcagctt	cccaagccgc	gcgactccat	gctggggagc	4140
cgtggggaga	tcacgcaccc	ctttgtggag	gtggagatca	ttgggctccc	tgtggactgc	4200
agcagggagc	agaccgcgt	ggtggacgac	aacgggttca	accccacctg	ggaggagacc	4260
ctggttttca	tgggtgcacat	gccggagatc	gcgctggtcc	gcttcctcgt	ctgggaccac	4320
gatcccatcg	ggcgtgactt	cattggccag	aggacgctgg	ccttcagcag	catgatgcca	4380

```

ggctacagac acgtgtacct agaagggatg gaagaggcct ccatcttcgt gcatgtggct 4440
gtcagtgaca tcagcggtaa ggtcaagcag gctctgggcc taaaaggcct cttcctccga 4500
ggcccaaagc ccggctcgct ggacagtcac gctgctgggc ggcccccgcc ccggccctcc 4560
gttagccagc ggaccctgcg gcgcacggcc agcgccccga ccaagagcca gaagccgggc 4620
cgcaggggct tccccgagct ggtcctgggt acacgggaca caggctccaa gggggtggca 4680
gacgatgtgg tgcccccgcc gcccgacct gctccggaag cccagccca ggaggggcc 4740
ggcagcgcca gccccgagg taaggcgcca gctgcggtgg cagagaagag ccctgtgcga 4800
gtgcggccct gcgtgtctg gacggccccc ggctgctgg gatggccgcc acatgcatga 4860
agtgtgtggt gggatcctgc gccggcgtag acaccggggg cctgcagagg gagcggccac 4920
ccagcccggg gcctgcaagc aggcaggcag ccattcgcca gcagccccgg gcccggtctg 4980
actactggg ggcacctgc tgtggcctgg accctcacgc tatcccgggg agaagcagag 5040
aggcccccaa gggtcctggg gcctggaggc aggggtccagg cggtagcggc tccatgtcct 5100
cggactccag cagcccagac agccccggca tccccgaaag gtccccccgc tggctgagg 5160
gtgcctcag gcaaccgggg gccctgcagg gagagatgag tgccttgttt gctcaaaagc 5220
tggaggagat caggagtaaa tccccatgt tctcccgcca caccgcccc ctctccacgc 5280
agcgccact cccccactg tgcagcctgg aaaccatgc tgaggagccc gccccaggcc 5340
ctggtcccc gccaccagc gctgtcccca ccagctcttc tcagggacgg ccccatacc 5400
ccacaggacc cggagccaat gtggcaagcc ccctagagga cactgaggag ccccgagaca 5460
gcaggcctcg gccgtgcaac ggcgagggcg ccggcggggc atacgagagg gccccggcca 5520
gccagacgga cggcaggagc cagccccgga ccctgggcca cctgccctg attagaagg 5580
tgaagagtga ggggcagggtg cccacggagc ccctgggagg gtggcgggcc ctggccgctc 5640
cctttccagc tctgcccgtg tactccgatg ccacgggcag tgaccgctg tggcagcggc 5700
tggagccatg tggccaaccga gacagcgttt cctcctctc cagcatgtca tccagcgaca 5760
ctgtcattga cctctccctg cccagcctgg gcctgggccc cagccgtgag aacctcgctg 5820
gagccacat gggacgcctg cccccaggc cccactcggc ttcggctgcc cggccagacc 5880
tgccacctgt gaccaagagc aaatccaacc ccaaccttc ggctacaggc cagcggcctc 5940
ccatacctga cgaactgcag cccaggctcc tggccccaag gatggctggc ctccccctcc 6000
ggcctccctg gggctgcctt tccctggtgg gcgtgcagga ctgccccgtg gctgccaaat 6060
ccaagagcct gggcgacctc actgctgatg actttgccc tagctttgag ggcggctccc 6120
gcagactgag ccacagcctg ggcctcccgg gagggacacg gcgggtgtcg gggccagggg 6180
tgagacggga caccctgaca gagcagctgc gctggctcac tgtcttcag caggcaggag 6240
acatcacgtc acccaccagc ctgggccccg ctggggaggg ggtggcaggg ggccttggtt 6300
ttgtgcggcg ctctcctcc cgcagccaca gcgcgtgcg tgccattgcc agccgggccc 6360
gccaggccca ggagcggcag cagagactgc agggcctggg ccggcaggga cccccgaag 6420
aggagcgggg cacccccag ggcgctgct ccgtgggcca cgagggcagt gtggatgcac 6480
cagcaccctc caagggagcc ctcgggccag catccgggc tgctgaaaac ctggtcctgc 6540
tccgctctg a 6551

```